Zero-Sum Average Cost Semi-Markov Games with Weakly Continuous Transition Probabilities and a Minimax Semi-Markov Inventory Problem

Óscar Vega-Amaya[1] · Fernando Luque-Vásquez[1] · Mauricio Castro-Enríquez[1]

Abstract
Under the framework given by a growth condition, a Lyapunov property and some continuity assumptions, the present work shows the existence of lower semicontinuous solutions to the Shapley equation for zero-sum semi-Markov games with Borel spaces, weakly continuous transition probabilities and possible unbounded payoff. It is also shown the existence of stationary optimal strategies for the minimizing player and stationary ε-optimal strategies for the maximizing player. These results are proved using a fixed-point approach. Moreover, it is shown the existence of a deterministic stationary minimax strategy for a minimax semi-Markov inventory problem under mild assumptions on the demand distribution.

Keywords Semi-Markov games · Average payoff · Lyapunov conditions · Shapley equation · Fixed-point approach

Mathematics Subject Classification 91A15 · 91A50

1 Introduction

This paper deals with zero-sum (ratio) average payoff semi-Markov games with Borel spaces, weakly continuous transition probabilities and unbounded lower semicontinuous payoff function. It is shown the existence of solutions to the Shapley equation and the existence of an optimal stationary strategy for the minimizing player and an ε-optimal stationary strategy for the maximizing one assuming that the game model satisfies a growth and Lyapunov conditions besides some continuity properties. The framework settled by these conditions was already used in several previous works dealing with Markov and semi-Markov

✉ Ó. Vega-Amaya
oscar.vega@unison.mx

F. Luque-Vásquez
fernando.luque@unison.mx

M. Castro-Enríquez
mauricio.castro@unison.mx

1 Departamento de Matemáticas, Universidad de Sonora, Hermosillo, Mexico

decision processes (for strongly continuous transition probabilities–in the control variable–see references [4, 6, 7, 21]; for weakly continuous transition probabilities–in the state-action pair–see references [9, 12, 13, 16, 24]) as well as with Markov and semi-Markov games (for strongly continuous transition probabilities see [8, 22]; for the weakly continuous case see [10, 11]). See also references [2, 17] for applications in communications systems. In fact, the present work extends the fixed-point approach of references [21, 24] to zero-sum average semi-Markov games with weakly transition probabilities and provides an application to a minimax semi-Markov inventory problem under fairly weak assumptions on the demand distributions. The readers are referred to reference [15] for a review on zero-sum stochastic games in discrete time. It is worth mentioning that zero-sum average payoff semi-Markov games were seemingly studied first by Tanaka and Wakuta [20],[1] who considered compact state and action spaces and assumed that the payoff function is continuous (thus, bounded), the transition law is weakly continuous among other conditions.

Concerning reference [10] some comments are in order. Jaśkiewicz [10] shows similar results to the present work under similar conditions but there are, of course, some important differences. To begin, to take advantage of the contraction property implied by the Lyapunov condition, she directly "smooths" functions–that are not lower semicontinuous–by taking liminf pointwise, which makes the proof of the validity of the Shapley equation somewhat technically involved. The present work gives a simpler proof of this result by using lower semicontinuous envelopes of functions and defining a suitable contraction operator. Moreover, the Lyapunov conditions of the present paper are weaker than those used in [10]. A second difference regards with the regularity property of the controlled processes, which is guaranteed in [10] by imposing essentially the standard condition on the holding time distribution (see, for instance, [19, Prop. 5.1(a), p. 88.]). Recall that the regularity property states that the involved stochastic processes experience finitely many transitions in bounded time intervals. This property, together the Lyapunov condition, plays an important role to guarantee that the ratio average payoff is well defined and finite-valued and also to show that the Shapley equation yields the existence of optimal or almost optimal stationary strategies for the players. These two latter facts are given by granted and not discussed explicitly by Jaśkiewicz [10]. In contrast, the present work does not impose any probabilistic condition additional to the Lyapunov condition; instead, in order to ensure that the ratio average payoff criterion is well defined, it is assumed that the admissible action sets for both players are compact sets. (Jaśkiewicz [10] supposes that the admissible action sets for one player are compact sets, while for the other player the admissible action sets are complete spaces.) In fact, the mentioned Lyapunov condition implies that the semi-Markov processes induced by stationary policies are regular (see [23, Theorem 18.3.4 and Remark 18.3.6.]). On the other hand, Jaśkiewicz [10] illustrates her results with a minimax Markov inventory problem with a finite number of possible distributions for the random demands, while the present work considers a minimax semi-Markov inventory problem assuming only that the first and second moments belong to bounded intervals. For additional results on minimax (or robust) control problems see references [3, 14, 16].

The remainder of the paper is organized as follows. Section 2 introduces the zero-sum semi-Markov game and the (ratio) average payoff performance index besides some standard concepts and notation. Section 3 states the assumptions and the main result (Theorem 3.4); its proof is given in Sect. 5. Section 4 shows the existence of a deterministic stationary minimax policy for a semi-Markov minimax inventory problem (see Theorem 4.2); the proof is given in Sect. 5.

---

[1]The authors thank to the referee for bringing their attention to this paper.

2 Zero-Sum Average Payoff Semi-Markov Games

The following standard concepts and notation are used throughout the paper. For a Borel space $\mathcal{S}$–that is, a subset of a complete and separable metric space–$\mathcal{B}(\mathcal{S})$ denotes the Borel σ-algebra; any statement about "measurability" should be meant as measurability with respect to $\mathcal{B}(\mathcal{S})$. The family of probability measures on $\mathcal{S}$ is denoted by $\mathbb{P}(\mathcal{S})$. Given two Borel spaces $\mathcal{S}$ and $\mathcal{S}'$, a *kernel* $K(\cdot|\cdot)$ on $\mathcal{S}$ given $\mathcal{S}'$ is a mapping such that $K(\cdot|s')$ is a measure on $\mathcal{S}$ for each $s' \in \mathcal{S}'$, and $K(B|\cdot)$ is a measurable function on $\mathcal{S}'$ for each subset $B \in \mathcal{B}(\mathcal{S})$. The kernel $K(\cdot|\cdot)$ is called *stochastic kernel* if $K(\cdot|s') \in \mathbb{P}(\mathcal{S})$ for all $s' \in \mathcal{S}'$.

Let $K(\cdot|\cdot)$ be a kernel on $\mathcal{S}$ given $\mathcal{S}'$. For an arbitrary measurable function $u : \mathcal{S} \to \mathbb{R}$, let

$$Ku(s') := \int_{\mathcal{S}} u(s)K(ds|s'), \quad s' \in \mathcal{S}',$$

whenever the integral is well defined. Similarly, for a measure $\nu \in \mathbb{P}(\mathcal{S})$, set

$$\nu(u) := \int_{\mathcal{S}} u(s)\nu(ds).$$

Given a measurable function $w : \mathcal{S} \to [1, \infty)$, $B_w(\mathcal{S})$ stands for the family of functions u on $\mathcal{S}$ with finite w-norm, which is defined as

$$||u||_w := \sup_{s \in \mathcal{S}} \frac{|u(s)|}{w(s)}.$$

By $L_w(\mathcal{S})$ denote the class of lower semicontinuous functions belonging to $B_w(\mathcal{S})$. The normed space $(B_w(\mathcal{S}), || \cdot ||_w)$ is a Banach space while $(L_w(\mathcal{S}), d_w)$ is a complete metric space, where d_w is the metric induced by the w-norm. The space of continuous bounded functions defined on $\mathcal{S}$ is denoted as $\mathcal{C}_b(\mathcal{S})$.

The subset of nonnegative real numbers is denoted by $\mathbb{R}_+$ and the positive (nonnegative, resp.) integers subset by $\mathbb{N}$ ($\mathbb{N}_0$, resp.).

The game model. We are interested in zero-sum semi-Markov games with the (ratio) average cost criterion given below (6). This kind of games is specified by a semi-Markov game model given by the collection

$$(\mathbf{X}, \mathbf{A}, \mathbf{B}, \mathbb{K}_\mathbf{A}, \mathbb{K}_\mathbf{B}, q, c), \tag{1}$$

where the Borel spaces $\mathbf{X}, \mathbf{A}, \mathbf{B}$ denote the *state space* of the game, and the action or control subsets for player 1 and player 2, respectively. The *constraint sets* $\mathbb{K}_\mathbf{A}$ and $\mathbb{K}_\mathbf{B}$ belongs to $\mathcal{B}(\mathbf{X} \times \mathbf{A})$ and $\mathcal{B}(\mathbf{X} \times \mathbf{B})$, respectively. The *x-sections*

$$A(x) := \{a \in \mathbf{A} : (x, a) \in \mathbb{K}_\mathbf{A}\}$$

$$B(x) := \{b \in \mathbf{B} : (x, b) \in \mathbb{K}_\mathbf{B}\}$$

stand for the *admissible action* or *control subsets* for players 1 and 2, respectively, when the game is in state $x \in \mathbf{X}$. The set

$$\mathbb{K} := \{(x, a, b) : x \in \mathbf{X}, a \in A(x), b \in B(x)\}$$

is a Borel subset of the Cartesian product $\mathbf{X} \times \mathbf{A} \times \mathbf{B}$ (see [18, Lemma 1.1]). The stochastic kernel $q(\cdot|\cdot, \cdot, \cdot)$ on $\mathbf{X} \times \mathbb{R}_+$ given $\mathbb{K}$ is the *transition law* of the game. Finally, the measurable function $c : \mathbb{K} \times \mathbb{R}_+ \to \mathbb{R}$ is the *payoff* function of the game.

The game is played over an infinite horizon as follows: at time $t = 0$, both players observe the game in some state, say, $x_0 = x \in \mathbf{X}$, and independently choose admissible controls $a_0 = a \in A(x)$ and $b_0 = b \in B(x)$. Then, the game remains in state $x_0 = x$ for a nonnegative random time Δ_1 and, at this time, it moves to a new state $x_1 = x' \in \mathbf{X}$ according to the probability measure $q(\cdot|x, a, b)$, that is,

$$q(B \times D|x, a, b) = \Pr[(x_1, \Delta_1) \in B \times D|x_0 = x, a_0 = a, b_0 = b], \tag{2}$$

for $B \in \mathcal{B}(\mathbf{X})$, $D \in \mathcal{B}(\mathbb{R}_+)$. Immediately after the transition occurs, player 1 pays the amount $c(x, a, b, \Delta_1)$ to player 2, and they choose new controls, say, $a_1 = a' \in A(x')$ and $b_1 = b' \in B(x')$, and the above process repeats over and over again.

This procedure engenders a stochastic process $\{(x_n, a_n, b_n, \Delta_{n+1})\}$, where, for each $n \in \mathbb{N}_0$, x_n is the state of the game, a_n and b_n are the control variables for players 1 and 2, respectively, and Δ_{n+1} is the time the game spends in state x_n; thus, the random time Δ_{n+1} is called *holding* or *sojourn time* at state x_n. Note that the random variable

$$T_n := T_{n-1} + \Delta_n, \ n \in \mathbb{N}, \ \text{ and } \ T_0 := 0,$$

is the time of the *n*th jump of the game. Thus, if $x_n = x$, $a_n = a$ and $b_n = b$, according to (2), the conditional marginal

$$Q(B|x, a, b) := q(B \times \mathbb{R}_+|x, a, b) \tag{3}$$

rules the state toward the game moves in the next transition irrespective of the time it takes to occur. Similarly, the conditional marginal

$$F(D|x, a, b) := q(\mathbf{X} \times D|x, a, b), \ D \in \mathcal{B}(\mathbb{R}_+),$$

rules the time at which the next transition happens irrespective the state toward which the system moves out; thus, it is called the *holding time distribution*. Then,

$$\tau(x, a, b) := \int_{\mathbb{R}^+} t F(dt|x, a, b) \tag{4}$$

is the *mean holding* or *sojourn time*, while

$$C(x, a, b) := \int_{\mathbb{R}^+} c(x, a, b, t) F(dt|x, a, b) \tag{5}$$

is the *mean payoff* that player 1 makes to player 2.

Strategies. Let $H_0 := \mathbf{X}$ and $H_n := \mathbb{K} \times \mathbb{R}_+ \times H_{n-1}$ for $n \in \mathbb{N}$. Thus, for $n \in \mathbb{N}$, each element

$$h_n = (x_0, a_0, b_0, \Delta_1, \ldots, x_{n-1}, a_{n-1}, b_{n-1}, \Delta_n, x_n) \in H_n$$

is the history of the game up to the *n*th transition, which occurs in the time T_n. A *strategy* for player 1 is a sequence $\pi^1 = \{\pi_n^1\}$ of stochastic kernels on $\mathbf{A}$ given H_n that satisfy the constraint

$$\pi_n^1(A(x_n)|h_n) = 1 \ \forall h_n \in H_n, n \in \mathbb{N}_0.$$

The class of all strategies for player 1 is denoted by Π^1. Now, for each $x \in \mathbf{X}$, let $\mathbb{A}(x) := \mathbb{P}(A(x))$ and denote by Φ^1 the class of stochastic kernels on $\mathbf{A}$ given $\mathbf{X}$ such that $\phi^1(\cdot|x) \in \mathbb{A}(x)$ for each $x \in \mathbf{X}$. A policy $\pi^1 = \{\pi_n^1\}$ is called *stationary* if

$$\pi_n^1(\cdot|h_n) = \phi^1(\cdot|x_n) \quad \forall h_n \in H_n, n \in \mathbb{N}_0$$

for some stochastic kernel $\phi^1 \in \Phi$. In this case, as usual, strategy $\pi^1 = \{\pi_n^1\}$ is identified with the stochastic kernel ϕ^1 and the class of all stationary strategies with Φ^1. A stationary policy ϕ^1 for player 1 is called *deterministic stationary* if there exists a measurable function $f : \mathbf{X} \to \mathbf{A}$ satisfying that $f(x) \in A(x)$ for all $x \in \mathbf{X}$ and such that $\phi^1(\cdot|x)$ is concentrated at $f(x)$ for each $x \in \mathbf{X}$; the set of deterministic stationary policies for player 1 is denoted by $\mathbb{F}^1$. The sets of strategies Π^2, Φ^2 and $\mathbb{F}^2$ for player 2 are defined similarly but considering $B(x)$ and $\mathbb{B}(x)$ in lieu of $A(x)$ and $\mathbb{A}(x)$, respectively.

Throughout of the remainder of the present work is used the following standard notation: for a measurable function u on $\mathbb{K}$, $x \in \mathbf{X}$ and probability measures $\gamma^1 \in \mathbb{A}(x)$, $\gamma^2 \in \mathbb{B}(x)$, let

$$u_{\gamma^1, \gamma^2}(x) := \int_{B(x)} \int_{A(x)} u(x, a, b) \gamma^1(da) \gamma^2(db).$$

Similarly, for a stationary strategy pair $(\phi^1, \phi^2) \in \Phi^1 \times \Phi^2$ and $x \in \mathbf{X}$, set

$$u_{\phi^1, \phi^2}(x) := \int_{B(x)} \int_{A(x)} u(x, a, b) \phi^1(da|x) \phi^2(db|x).$$

Thus, in particular,

$$C_{\phi^1, \phi^2}(x) = \int_{B(x)} \int_{A(x)} C(x, a, b) \phi^1(da|x) \phi^2(db|x),$$

$$\tau_{\phi^1, \phi^2}(x) = \int_{B(x)} \int_{A(x)} \tau(x, a, b) \phi^1(da|x) \phi^2(db|x),$$

and also

$$Q_{\phi^1, \phi^2}(\cdot|x) = \int_{B(x)} \int_{A(x)} Q(\cdot|x, a, b) \phi^1(da|x) \phi^2(db|x).$$

The average payoff performance index. Let $\Omega := (\mathbb{K} \times \mathbb{R}_+)^\infty$ and $\mathcal{F}$ the corresponding product σ-algebra. For each strategy pair $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and probability measure ν on $\mathbf{X}$ there exist a probability measure $P_\nu^{(\pi^1, \pi^2)}$ and a stochastic process $\{(x_n, a_n, b_n, \Delta_{n+1})\}$ defined on the sample space $(\Omega, \mathcal{F})$ with the following properties:

(*i*) $P_\nu^{(\pi^1, \pi^2)}[x_0 \in B] = \nu(B)$;

(*ii*) $P_\nu^{(\pi^1, \pi^2)}[(a_n, b_n) \in C_1 \times C_2 | h_n] = \pi_n^1(C_1 | h_n) \pi_n^2(C_2 | h_n)$;

(*iii*) $P_\nu^{(\pi^1, \pi^2)}[(x_{n+1}, \Delta_{n+1}) \in B \times D | h_n, a_n, b_n] = q(B \times D | x_n, a_n, b_n)$;

for all $B \in \mathcal{B}(\mathbf{X})$, $C_1 \in \mathcal{B}(\mathbf{A})$, $C_2 \in \mathcal{B}(\mathbf{B})$, $D \in \mathcal{B}(\mathbb{R}_+)$, $h_n \in H_n$, $n \in \mathbb{N}$.

The expectation operator with respect to $P_\nu^{(\pi^1, \pi^2)}$ is denoted as $E_\nu^{(\pi^1, \pi^2)}$. According to property (*i*), the probability measure ν is called *initial distribution*. If the initial distribution ν is concentrated at some state $x \in \mathbf{X}$, we shall write $P_x^{(\pi^1, \pi^2)}$ and $E_x^{(\pi^1, \pi^2)}$ instead of $P_\nu^{(\pi^1, \pi^2)}$ and $E_\nu^{(\pi^1, \pi^2)}$, respectively.

If the players use a stationary strategy pair $(\phi^1, \phi^2) \in \Phi^1 \times \Phi^2$, by property *(iii)*, the state process $\{x_n\}$ is a Markov chain with one-step transition probability $Q_{\phi^1, \phi^2}(\cdot|\cdot)$. In this case, the n-step transition probability is denoted by $Q^n_{\phi^1, \phi^2}(\cdot|\cdot)$ for $n \in \mathbb{N}_0$, where $Q^0_{\phi^1, \phi^2}(\cdot|x)$ is the Dirac measure concentrated at $x \in \mathbf{X}$. Thus, for a measurable function $u : \mathbf{X} \to \mathbb{R}$,

$$Q^n_{\phi^1, \phi^2} u(x) = \int_{\mathbf{X}} u(y) Q^n_{\phi^1, \phi^2}(dy|x) = E_x^{(\phi^1, \phi^2)} u(x_n) \quad \forall x \in \mathbf{X},$$

whenever these quantities are well defined.

The (ratio) *expected average payoff* (EAP) for a strategy pair $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and initial state $x \in \mathbf{X}$ is defined as

$$J(\pi^1, \pi^2, x) := \limsup_{n \to \infty} \frac{E_x^{(\pi^1, \pi^2)} \sum_{i=0}^{n-1} c(x_i, a_i, b_i, \Delta_{i+1})}{E_x^{(\pi^1, \pi^2)} T_n}. \tag{6}$$

Since the equalities

$$E_x^{(\pi^1, \pi^2)} \Delta_{i+1} = E_x^{(\pi^1, \pi^2)} \tau(x_i, a_i, b_i), \tag{7}$$

$$E_x^{(\pi^1, \pi^2)} c(x_i, a_i, b_i, \Delta_{i+1}) = E_x^{(\pi^1, \pi^2)} C(x_i, a_i, b_i) \tag{8}$$

hold for all strategy pair $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, state $x \in \mathbf{X}$ and $i \in \mathbb{N}_0$, the performance index (6) can be re-written as

$$J(\pi^1, \pi^2, x) = \limsup_{n \to \infty} \frac{E_x^{(\pi^1, \pi^2)} \sum_{i=0}^{n-1} C(x_i, a_i, b_i)}{E_x^{(\pi^1, \pi^2)} \sum_{i=0}^{n-1} \tau(x_i, a_i, b_i)}. \tag{9}$$

Roughly speaking, the goal of player 1 (player 2, resp.) is to minimize (maximize, resp.) (9). This leads to the *lower* and *upper value* functions

$$L(\cdot) := \sup_{\pi^2 \in \Pi^2} \inf_{\pi^1 \in \Pi^1} J(\pi^1, \pi^2, \cdot) \quad \text{and} \quad U(\cdot) := \inf_{\pi^1 \in \Pi^1} \sup_{\pi^2 \in \Pi^2} J(\pi^1, \pi^2, \cdot),$$

respectively. Observe that, in general, $L(\cdot) \leq U(\cdot)$. Thus, if $L(\cdot) = U(\cdot)$, it is said that the game has a value and the common value of these functions, which is denoted as $V(\cdot)$, is called the *value of the game*.

If the game has value $V(\cdot)$, a policy $\pi^1_* \in \Pi^1$ is said to be *EAP-optimal* for player 1 if

$$\sup_{\pi^2 \in \Pi^2} J(\pi^1_*, \pi^2, \cdot) = V(\cdot);$$

similarly, a policy $\pi^2_* \in \Pi^2$ is said to be *EAP-optimal* for player 2 if

$$\inf_{\pi^1 \in \Pi^1} J(\pi^1, \pi^2_*, \cdot) = V(\cdot).$$

If π^i_* is EAP-optimal for player i $(i = 1, 2)$, then the pair (π^1_*, π^2_*) is called *EAP-optimal pair* or *saddle point*. Observe that (π^1_*, π^2_*) is an EAP-optimal pair if and only if

$$J(\pi^1_*, \pi^2, \cdot) \leq J(\pi^1_*, \pi^2_*, \cdot) \leq J(\pi^1, \pi^2_*, \cdot) \quad \forall \pi^1 \in \Pi^1, \pi^2 \in \Pi^2.$$

Remark 2.1 An important particular case of semi-Markov games are Markov games, which come out when the holding time distribution is concentrated at some positive constant c, say $c = 1$, that is

$$F(\{1\}|x, a, b) = 1 \ \forall (x, a, b) \in \mathbb{K}.$$

In this case,

$$P_x^{(\pi^1, \pi^2)}[T_n = n] = 1 \quad \forall x \in \mathbf{X}, (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2, n \in \mathbb{N},$$

and $\tau = 1$. Thus, the performance index (9) becomes

$$J(\pi^1, \pi^2, x) = \limsup_{n \to \infty} \frac{1}{n} E_x^{(\pi^1, \pi^2)} \sum_{i=0}^{n-1} C(x_i, a_i, b_i).$$

Remark 2.2 The performance index (9) depends on the kernel q only through its marginal distribution measures Q and F, which enter in the computation in a "decoupled" way. Thus, one can assume without loss of generality that

$$q(B \times D|x, a, b) = F(D|x, a, b)Q(B|x, a, b)$$

for all $(x, a, b) \in \mathbb{K}$, $B \in \mathcal{B}(\mathbf{X})$, $D \in \mathcal{B}(\mathbb{R}_+)$. A second consequence is that the mean holding time distribution F can be replaced by an exponential distribution with parameter τ^{-1}, that is, by the distribution

$$F'(D|x, a, b) := \tau^{-1}(x, a, b) \int_D e^{-\tau^{-1}(x,a,b)t} dt$$

under the assumption that $\tau > 0$. In other words, the semi-Markov game model (1) can be replaced without loss of generality by the model $(\mathbf{X}, \mathbf{A}, \mathbf{B}, \mathbb{K}_{\mathbf{A}}, \mathbb{K}_{\mathbf{B}}, q', C)$ where

$$q'(B \times D|x, a, b) := F'(D|x, a, b)Q(B|x, a, b),$$

Q is the transition kernel (3) and C is the cost function (8). It is worth mentioning that this could not be the case for a discounted performance index.

3 Solutions to the Shapley Equation and Optimal Stationary Strategies

The main result of the present work, Theorem 3.4 below, extends to semi-Markov games with weak continuous transition probabilities the analysis given in [24] for Markov decision processes with continuous transition probabilities. Specifically, it is shown the existence of lower semicontinuous solutions to the Shapley equation under Assumptions 1, 2 and 3 given below. As commented in the Introduction, the framework settled by these conditions has became quite standard for the study of average payoff optimization problems and used, for instance, for Markov and semi-Markov decision processes in [4, 6, 7, 9, 12, 13, 16, 21, 24], and for Markov and semi-Markov games in [10, 11, 15, 22]. See reference [5] for further comments and discussion on such kind of conditions.

Assumption 1 The following conditions hold for all $(x, a, b) \in \mathbb{K}$:

(a) $\tau(x, a, b) > 0$;

(b) there exist a measurable function $W \geq 1$ on $\mathbf{X}$ and a constant $k > 0$ such that

$$\max\{\tau(x, a, b), |C(x, a, b)|\} \leq kW(x).$$

Assumption 2 There exist a measurable nonnegative function s on $\mathbb{K}$, a nontrivial measure ν on $\mathbf{X}$, and a constant $\lambda \in (0, 1)$ such that the following properties hold:

(a) $\nu(W) < \infty$;

(b) $Q(B|x, a, b) \geq s(x, a, b)\nu(B)$ for all $(x, a, b) \in \mathbb{K}$;

(c) $QW(x, a, b)) \leq \lambda W(x) + \nu(W)s(x, a, b)$ for all $(x, a, b) \in \mathbb{K}$;

(d) $\nu(s_{\phi^1,\phi^2}) > 0$.

The key point is that Assumption 2 entails a contraction property. To see this, let

$$\widehat{Q}(B|x, a, b) := Q(B|x, a, b) - s(x, a, b)\nu(B), \quad (x, a, b) \in \mathbb{K}, B \in \mathcal{B}(\mathbf{X}).$$

Then, Assumption 2(b) implies that $\widehat{Q}$ is a nonnegative kernel, while Assumption 2(c) leads to the inequality

$$\sup_{(x,a,b)} \frac{|\widehat{Q}u(x, a, b)|}{W(x)} \leq \lambda ||u||_W \quad \forall u \in B_W(\mathbf{X}).$$

The results in the next proposition are proved in [21] using this contraction property.

Proposition 3.1 *Suppose that Assumption 2 holds and let $(\phi^1, \phi^2) \in \Phi^1 \times \Phi^2$ be an arbitrary stationary strategy pair. Then*:

(a) the transition probability Q_{ϕ^1,ϕ^2} is positive Harris recurrent; thus, it admits a unique invariant probability measure μ_{ϕ^1,ϕ^2} and ν is an irreducibility measure;

(b) $\mu_{\phi^1,\phi^2}(W)$ is finite and

$$1 \leq \mu_{\phi^1,\phi^2}(W) \leq \frac{1}{1 - \lambda} \frac{\nu(W)}{\nu(\mathbf{X})};$$

moreover,

$$\mu_{\phi^1,\phi^2}(s_{\phi^1,\phi^2}) \geq \theta := \frac{1 - \lambda}{\nu(W)};$$

(c) for any $u \in B_W(\mathbf{X})$, $\mu_{\phi^1,\phi^2}(|u|) < \infty$ and

$$\lim_{n\to\infty} \frac{1}{n} E_x^{(\phi^1,\phi^2)} \sum_{i=0}^{n-1} u(x_i) = \mu_{\phi^1,\phi^2}(u) \quad \forall x \in \mathbf{X};$$

hence,

$$\lim_{n\to\infty} \frac{1}{n} E_x^{(\phi^1,\phi^2)} u(x_n) = 0.$$

Remark 3.2 Thus, for each pair $(\phi^1, \phi^2) \in \Phi^1 \times \Phi^2$, Assumptions 1 and 2 imply that the constant

$$\rho(\phi^1, \phi^2) := \frac{\mu_{\phi^1,\phi^2}(C_{\phi^1,\phi^2})}{\mu_{\phi^1,\phi^2}(\tau_{\phi^1,\phi^2})}, \tag{10}$$

is well defined and finite, and also that the operator

$$T_{\phi^1,\phi^2} u := C_{\phi^1,\phi^2} - \rho(\phi^1, \phi^2)\tau_{\phi^1,\phi^2} + \widehat{Q}_{\phi^1,\phi^2} u$$

is a contraction map from $B_W(\mathbf{X})$ into itself with modulus λ. Since $(B_W(\mathbf{X}), ||\cdot||_W)$ is a Banach space, it follows that there exists a unique function $h_{\phi^1,\phi^2} \in B_W(\mathbf{X})$ such that

$$\begin{aligned} h_{\phi^1,\phi^2} &= T_{\phi^1,\phi^2} h_{\phi^1,\phi^2} \\ &= C_{\phi^1,\phi^2} - \rho(\phi^1, \phi^2)\tau_{\phi^1,\phi^2} + \widehat{Q}_{\phi^1,\phi^2} h_{\phi^1,\phi^2}. \end{aligned} \tag{11}$$

Now an integration of both sides of the above equation with respect to the probability measure μ_{ϕ^1,ϕ^2} leads to equality

$$0 = \nu(h_{\phi^1,\phi^2})\mu_{\phi^1,\phi^2}(s_{\phi^1,\phi^2}),$$

which implies, by Proposition 3.1(b), that $\nu(h_{\phi^1,\phi^2}) = 0$. Hence, h_{ϕ^1,ϕ^2} is the unique function on $B_W(\mathbf{X})$ that satisfies the (semi-Markov) Poisson equation

$$h_{\phi^1,\phi^2} = C_{\phi^1,\phi^2} - \rho(\phi^1, \phi^2)\tau_{\phi^1,\phi^2} + Q_{\phi^1,\phi^2} h_{\phi^1,\phi^2}$$

and the condition $\nu(h_{\phi^1,\phi^2}) = 0$. Then, by iterating this equation, one can see that

$$J(\phi^1, \phi^2, x) = \lim_{n\to\infty} \frac{E_x^{(\phi^1,\phi^2)} \sum_{i=0}^{n-1} C_{\phi^1,\phi^2}(x_i)}{E_x^{(\phi^1,\phi^2)} \sum_{i=0}^{n-1} \tau_{\phi^1,\phi^2}(x_i)} = \rho(\phi^1, \phi^2) \tag{12}$$

for all $x \in \mathbf{X}$.

It is worth mentioning that reference [21] proves Proposition 3.1 using similar arguments to those displayed above to show the existence of solutions to the Poisson equation (11).

Now, to pass from the Poisson equations to the Shapley equation, it is assumed that the model game satisfies the following compactness and continuity conditions.

Assumption 3 (a) C is lower semicontinuous on $\mathbb{K}$;
   (b) the mapping $x \to A(x)$ is lower semicontinuous and compact-valued;
   (c) the mapping $x \to B(x)$ is upper semicontinuous and compact valued;
   (d) τ is continuous;
   (e) the state transition law Q is weakly continuous on $\mathbb{K}$, that is, the mapping

$$(x, a, b) \to \int_{\mathbf{X}} u(y) Q(dy|x, a, b)$$

is continuous for all $u \in \mathcal{C}_b(\mathbf{X})$;
   (f) W and QW are continuous functions.

Notice that Remark 3.2 shows that the average payoff criterion is well defined and finite-valued whenever the players use stationary strategies. The next theorem extends this assertion to all admissible strategies.

Theorem 3.3 *Under Assumptions* 1, 2 *and* 3, *the performance criterion* (9) *is well defined and finite-valued.*

The next theorem states the main results of the present work.

Theorem 3.4 *If Assumptions* 1, 2, 3 *hold, then:*
  (*a*) *there exist* $h^* \in L_W(\mathbf{X})$ *and* $\rho^* \in \mathbb{R}$ *that satisfy the Shapley equation*

$$h^*(x) = \inf_{\gamma^1 \in \mathbb{A}(x)} \sup_{\gamma^2 \in \mathbb{B}(x)} [C_{\gamma^1,\gamma^2}(x) - \rho^* \tau_{\gamma^1,\gamma^2}(x) + Q_{\gamma^1,\gamma^2} h^*(x)]$$

$$= \sup_{\gamma^2 \in \mathbb{B}(x)} \inf_{\gamma^1 \in \mathbb{A}(x)} [C_{\gamma^1,\gamma^2}(x) - \rho^* \tau_{\gamma^1,\gamma^2}(x) + Q_{\gamma^1,\gamma^2} h^*(x)]$$

for all $x \in \mathbf{X}$;
  (*b*) *there exists* $\phi_*^1 \in \Phi^1$ *such that*

$$h^*(x) = \sup_{\gamma^2 \in \mathbb{B}(x)} [C_{\phi_*^1,\gamma^2}(x) - \rho^* \tau_{\phi_*^1,\gamma^2}(x) + Q_{\phi_*^1,\gamma^2} h^*(x)] \quad \forall x \in \mathbf{X};$$

  (*c*) *for each* $\varepsilon > 0$ *there exists* $\phi_\varepsilon^2 \in \Phi^2$ *such that*

$$h^*(x) \leq \inf_{\gamma^1 \in \mathbb{A}(x)} [C_{\gamma^1,\phi_\varepsilon^2}(x) - \rho^* \tau_{\gamma^1,\phi_\varepsilon^2}(x) + Q_{\gamma^1,\phi_\varepsilon^2} h^*(x)] + \varepsilon \quad \forall x \in \mathbf{X}.$$

  (*d*) *the constant* ρ^* *is the value of the game, the stationary policy* ϕ_*^1 *is optimal for player* 1; *for each* $\varepsilon > 0$, *the stationary policy* $\phi_{\varepsilon'}^2$ *is* ε-*optimal for player* 2 *with* $\varepsilon' = v(W)l\varepsilon/(1 - \lambda)$ *where* l *is a fixed constant* (*this is given below in Lemma* 5.3);
  (*e*) *moreover,*

$$\rho^* = \inf_{\phi^1 \in \Phi^1} \sup_{\phi^2 \in \Phi^2} \rho(\phi^1, \phi^2) = \sup_{\phi^2 \in \Phi^2} \inf_{\phi^1 \in \Phi^1} \rho(\phi^1, \phi^2),$$

The proof of Theorem 3.4 is given in Sect. 4.

Remark 3.5 Assumption 3(d) trivially holds for Markov games (see Remark 2.1). Moreover, as can be checked in the proof of Theorem 3.4, instead of the compactness property in Assumption 3(c), it suffices to assume the completeness of the sets $B(x)$, $x \in \mathbf{X}$. In fact, the compactness property is used only to prove the existence of a positive constant γ such that

$$\liminf_{m \to \infty} \frac{1}{m} \sum_{i=0}^{m-1} E_x^{(\pi^1, \pi^2)} \tau(x_i, a_i, b_i) \geq \gamma$$

for all initial state $x \in \mathbf{X}$ and strategy pair $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ (see Corollary 5.5 in Sect. 5). This latter inequality clearly holds for Markov games.

4 A Minimax Semi-Markov Inventory Problem

A controller seeks to minimize the cost of operating a single-item inventory system without backlog, for which the decision epochs form a nondecreasing stochastic processes $T_n, n \in \mathbb{N}_0$. The demand process $\{w_n\}$ is formed by independent nonnegative random variables with w_n being the quantity of product demanded between the decision epochs T_n and T_{n+1}. The controller knows that the expected demands lie in a bounded interval and that its second moments are bounded above, but she/he does not know the demand distribution themselves. To state these assumptions formally, for $p \in \mathbb{P}([0, \infty))$, let

$$\mu_p := \int_0^\infty s p(ds) \quad \text{and} \quad s_p^2 := \int_0^\infty s^2 p(ds).$$

Assumption 4 The demand distributions belong to the class

$$\mathbf{B} := \{b \in \mathbb{P}(\mathbb{R}_+) : z_* \leq \mu_b \leq z^*, s_b^2 \leq s^*\},$$

where z^*, z_* and s^* are constants such that $z^* > z_* > 0$ and $s^* \geq 0$.

In what follows it is assumed that $\mathbf{B}$ is metrized with any metric compatible with the topology of weak convergence of probability measures.

The inventory evolves as follows: at the *nth* decision epoch, which is given by the time T_n, the controller observes the inventory level $x_n = x \in \mathbf{X} := \mathbb{R}_+$ and orders a product quantity $a_n = a \in \mathbf{A} := [0, \widehat{a}]$ for facing the (nonnegative) product demand $w_n = w$ accumulated between times T_n and T_{n+1}; the constant $\widehat{a}$ is positive and it is assumed that the replenishing quantity a_n is immediately supplied. Thus, the controller incurs in the cost

$$c_1(1 - \delta_0(a)) + c_2 a + c_3(x + a) + c_4(w - x - a)^+$$

where $\delta_0(a)$ is the delta of Kronecker and the nonnegative constants $c_i, i = 1, \ldots, 4$, stand for the cost for putting an order, the production cost, the holding cost and the penalizating cost for unmet demand, respectively, per unit of product.

Then, the expected payoff (or one step cost) is

$$C(x, a, b) := c_1(1 - \delta_0(a)) + c_2 a + c_3(x + a) + c_4 E_b(w - x - a)^+ \tag{13}$$

for $(x, a, b) \in \mathbb{K}$, with $A(x) := \mathbf{A}$ and $B(x) := \mathbf{B}$ for all $x \in \mathbf{X}$, and where E_b denotes the expectation with respect to the distribution $b \in \mathbf{B}$. Notice that $\mathbb{K} = \mathbf{X} \times \mathbf{A} \times \mathbf{B}$.

At time T_{n+1} the inventory changes according to the recursive equation

$$x_{n+1} = (x_n + a_n - w_n)^+, n \in \mathbb{N}_0, \tag{14}$$

where $r^+ := max(r, 0)$ for $r \in \mathbb{R}$, which leads to the stochastic kernel

$$Q(B|x, a, b) := b([w \geq 0 : (x + a - w)^+ \in B]), \quad B \in \mathcal{B}(\mathbf{X}), (x, a, b) \in \mathbb{K}. \tag{15}$$

Since the distribution of the product demand is unknown, the controller wants to hedge himself against the worse possible scenario; then, she/he approaches the problem as a game against "nature", who choses the distribution $b_n \in \mathbf{B}$ for the product demand w_n. Thus,

the controller goal is to find a *minimax policy*, that is, a policy $\pi_*^1 \in \Pi^1$ that satisfies the condition

$$\sup_{\pi^2 \in \Pi^2} J(\pi_*^1, \pi^2, x) = \inf_{\pi^1 \in \Pi^1} \sup_{\pi^2 \in \Pi^2} J(\pi^1, \pi^2, x) \quad \forall x \in \mathbf{X}. \tag{16}$$

Notice that a minimax policy is no other than an optimal policy for the controller when the inventory system is seen as a game against nature and this game has a value.

Jaśkiewicz [10] considers a minimax Markov inventory problem, that is, a minimax inventory problem for which $T_n = n$ for all $n \in \mathbb{N}$; moreover, she assumes that the set **B** of possible distributions for the demand is finite. Reference [16] also follows the minimax approach to study a class of semi-Markov inventory system but assuming that the holding time distribution depends on an unknown parameter and that the demand distribution is completely known.

The main result of this section, Theorem 4.2 given below, shows the existence of a *deterministic stationary* minimax policy for the inventory system (13)-(15) under Assumption 4 and Assumptions 5 and 6, which are given next. Assumption 5 is also related to the distributions of the product demands.

Assumption 5 The inequality $z_* > \widehat{a}$ holds.

It is shown in Lemma 6.1, Sect. 6, that Assumptions 4 and 5 imply that there exists a constant $r > 0$ such that

$$\lambda := \sup_{b \in \mathbf{B}} \Phi_b(r) < 1, \tag{17}$$

where

$$\Phi_p(t) := \int_{\mathbb{R}_+} e^{t(\widehat{a}-s)} p(ds), \quad t \geq 0, \ p \in \mathbb{P}(\mathbb{R}_+). \tag{18}$$

This inequality plays a key role in proving that the inventory system satisfies Assumption 2. In fact, Lemma 6.2 shows that the constant λ in (17), the functions

$$W(x) := \exp(rx) \quad \text{and} \quad s(x, a, b) := b((x + a, \infty)), \ (x, a, b) \in \mathbb{K}, \tag{19}$$

and the Dirac measure at zero $\nu(\cdot)$ satisfy Assumption 2.

The next condition concerns with the mean holding time distribution. In order to be specific, it is assumed that it has an exponential density where the parameter κ^{-1} is a positive continuous function dominated by the function W; such condition is formally stated below. However, it should be noted that it can be considered any other class of distributions as long as Assumption 1 holds (see Remark 2.2).

Assumption 6 The mean holding time distribution is given as

$$F([0, t]|x, a, b) = \kappa^{-1}(x, a, b) \int_0^t e^{-\kappa^{-1}(x,a,b)s} ds$$

for all $t \geq 0$ and $(x, a, b) \in \mathbb{K}$, where $\kappa : \mathbb{K} \to (0, \infty)$ is a continuous function such that $\kappa(\cdot, \cdot, \cdot) \leq k_1 W(\cdot)$ for some positive constant k_1.

Thus, according to Remark 2.2, the semi-Markov kernel can be taken without loss of generality as

$$q(B \times [0, t]|x, a, b) = [1 - e^{-\kappa^{-1}(x,a,b)t}]b([w \geq 0 : (x + a - w)^+ \in B])$$

for all $t \geq 0$, $(x, a, b) \in \mathbb{K}$ and $B \in \mathcal{B}(\mathbb{R}_+)$.

Remark 4.1 Observe that Assumption 4 implies the inequality

$$0 \leq C(x, a, b) \leq M + c_3 x \quad \forall (x, a, b) \in \mathbb{K},$$

where $M := c_1 + (c_2 + c_3)\widehat{a} + c_4 z^*$. For the other hand, Assumption 6 implies that

$$0 < \tau(x, a, b) = \kappa(x, a, b) \leq k_1 W(x) \quad \forall (x, a, b) \in \mathbb{K}.$$

Thus, one can choose a constant $k \geq k_1 > 0$ such that

$$\max\{\tau(x, a, b), |C(x, a, b)|\} \leq k W(x),$$

for all $(x, a, b) \in \mathbb{K}$. Notice that this latter inequality shows that Assumption 1(b) holds.

Next, it is the main result of this section.

Theorem 4.2 *Suppose that Assumptions 4, 5 and 6 hold. Then there exists a function $h^* \in L_W(\mathbf{X})$, a constant ρ^* and a deterministic stationary policy $f_*^1 \in \mathbb{F}^1$ for player 1 such that the equalities*

$$h^*(x) = \inf_{a \in \mathbf{A}} \sup_{b \in \mathbf{B}} [C(x, a, b) - \rho^* \tau(x, a, b) + \int_{\mathbf{X}} h^*(y) Q(dy|x, a, b)]$$

$$= \sup_{b \in \mathbf{B}} [C(x, f_*^1(x), b) - \rho^* \tau(x, f_*^1(x), b) + \int_{\mathbf{X}} h^*(y) Q(dy|x, f_*^1(x), b)]$$

hold for all $x \in \mathbf{X}$. Moreover, f_*^1 is a minimax policy and

$$\rho^* = \sup_{\pi^2 \in \Pi^2} J(f_*^1, \pi^2, x) = \inf_{\pi^1 \in \Pi^1} \sup_{\pi^2 \in \Pi^2} J(\pi^1, \pi^2, x) \quad \forall x \in \mathbf{X}.$$

5 Proof of Theorems 3.3 and 3.4

The proof relies on a number of preliminary results, which are collected in several propositions and lemmas. In Proposition 5.1, it is used the concept of lower semicontinuous envelope of functions, which is introduced below together with some related properties.

Let $(\mathcal{S}, d)$ be a metric space. For each function $u : \mathcal{S} \to \mathbb{R}$ define

$$u^e(s) := \sup_{r > 0} \inf_{s' \in B_r(s)} u(s'),$$

where $B_r(s)$ is the ball with center at $s \in \mathcal{S}$ and radius $r > 0$. Function u^e is the largest lower semicontinuous function dominated by u, that is: *(i)* u^e is lower semicontinuous; *(ii)* $u \geq u^e$; *(iii) if v is lower semicontinuous and $u \geq v$, then $u^e \geq v$. Thus, u^e is called the *lower*

semicontinuous envelope of the function u. Clearly, u is lower semicontinuous if and only if $u = u^e$; moreover, if $u \geq v$, then $u^e \geq v^e$. Additionally, if $w : S \to [1, \infty)$ is a continuous function, then

$$||u^e - v^e||_w \leq ||u - v||_w \quad \forall u, v \in B_w(S).$$

In what follows is assumed that Assumptions 1, 2, 3 hold.

Proposition 5.1 *Let* $S := s^e$. *Then*:

    (a) $Q(B|x, a, b) \geq S(x, a, b)\nu(B)$ *for all* $(x, a, b) \in \mathbb{K}$;

    (b) $QW(x, a, b) \leq \lambda W(x) + \nu(W)S(x, a, b)$ *for all* $(x, a, b) \in \mathbb{K}$;

    (c) $\mu_{\phi^1, \phi^2}(S_{\phi^1, \phi^2}) \geq \theta > 0$ *for all pair* $(\phi^1, \phi^2) \in \Phi^1 \times \Phi^2$.

Proof Property (a) follows directly from the definition of lower semicontinuous envelope. To prove (b), note that $\nu(W)s \geq QW - \lambda W$ and recall that QW and W are continuous functions; then

$$\nu(W)S = (\nu(W)s)^e \geq (QW - \lambda W)^e = QW - \lambda W,$$

which proves (b). Now, note that part (b) yields for each pair of stationary strategies (ϕ^1, ϕ^2) the inequality

$$Q_{\phi^1, \phi^2} W \leq \lambda W + \nu(W)S_{\phi^1, \phi^2}.$$

Thus, the inequality resulting by integrating with respect to μ_{ϕ^1, ϕ^2} at both sides implies that

$$\mu_{\phi^1, \phi^2}(S_{\phi^1, \phi^2}) \geq (1 - \lambda)\frac{\mu_{\phi^1, \phi^2}(W)}{\nu(W)} \geq \frac{1 - \lambda}{\nu(W)} = \theta,$$

which proves part (c). $\square$

Proposition 5.2 *For all strategy pair* $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $x \in \mathbf{X}$ *and* $u \in B_W(\mathbf{X})$:

$$\lim_{n \to \infty} \frac{1}{n} E_x^{(\pi^1, \pi^2)} u(x_n) = 0.$$

Proof This result follows directly after noting that Assumption 1(b) and Assumption 2(c) imply that

$$1 \leq E_x^{(\pi^1, \pi^2)} W(x_n) \leq \lambda^n W(x) + \frac{\nu(W)}{(1 - \lambda)\nu(X)} \quad \forall n \in \mathbb{N}_0. \tag{20}$$

$\square$

Lemma 5.3 *There exists a constant* $l > 0$ *such that*

$$E_x^{(\pi^1, \pi^2)} \tau(x_k, a_k, b_k) \geq l E_x^{(\pi^1, \pi^2)} s(x_{k-1}, a_{k-1}, b_{k-1})$$

for all $x \in \mathbf{X}$, $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $k \in \mathbb{N}$.

Proof Fix $x \in \mathbf{X}$, $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $k \in \mathbb{N}_0$. Now, observe that

$$\tau^*(x) := \min_{(a,b) \in A(x) \times B(x)} \tau(x, a, b) > 0.$$

Next consider the subsets

$$\mathbf{X}_n := \{x \in \mathbf{X} : \tau^*(x) \geq 1/n\}$$

and notice that $\mathbf{X}_n \uparrow \mathbf{X}$. Thus, for some $N \in \mathbb{N}$, $\nu(\mathbf{X}_N) > 0$ since $\nu(\mathbf{X}_n) \uparrow \nu(\mathbf{X}) > 0$. Then,

$$
\begin{aligned}
E_x^{(\pi^1,\pi^2)} \tau(x_k, a_k, b_k) &= E_x^{(\pi^1,\pi^2)} E_x^{(\pi^1,\pi^2)}[\tau(x_k, a_k, b_k)|h_{k-1}, a_{k-1}, b_{k-1}] \\
&= E_x^{(\pi^1,\pi^2)} \int_{\mathbf{X}} \int_{\mathbf{A}} \int_{\mathbf{B}} \tau(y, a, b) \pi_k^1(da|h_k) \pi_k^2(db|h_k) Q(dy|x_{k-1}, a_{k-1}, b_{k-1}) \\
&\geq E_x^{(\pi^1,\pi^2)} \int_{\mathbf{X}_N} \int_{\mathbf{A}} \int_{\mathbf{B}} \frac{1}{N} \pi_k^1(a|h_k) \pi_k^2(db|h_k) Q(dy|x_{k-1}, a_{k-1}, b_{k-1}) \\
&= \frac{1}{N} E_x^{(\pi^1,\pi^2)} Q(X_N|x_{k-1}, a_{k-1}, b_{k-1}) \\
&\geq \frac{1}{N} \nu(\mathbf{X}_N) E_x^{(\pi^1,\pi^2)} s(x_{k-1}, a_{k-1}, b_{k-1});
\end{aligned}
$$

which proves the desired result with $l := \nu(\mathbf{X}_N)/N$. $\qquad\square$

Lemma 5.4 *The following inequality holds for all $x \in \mathbf{X}$, $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$:*

$$\liminf_{m \to \infty} \frac{1}{m} E_x^{(\pi^1,\pi^2)} \sum_{k=0}^{m-1} s(x_k, a_k, b_k) \geq \frac{1-\lambda}{\nu(W)}.$$

Proof Let $x \in \mathbf{X}$, $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ be fixed but arbitrary. It follows from Assumption 2(c) that

$$
\begin{aligned}
E_x^{(\pi^1,\pi^2)} s(x_{k-1}, a_{k-1}, b_{k-1}) &\geq \frac{1}{\nu(W)} E_x^{(\pi^1,\pi^2)}[QW(x_{k-1}, a_{k-1}, b_{k-1}) - \lambda W(x_{k-1})] \\
&= \frac{1}{\nu(W)} E_x^{(\pi^1,\pi^2)}[W(x_k) - W(x_{k-1}) + (1-\lambda)W(x_{k-1})].
\end{aligned}
$$

Then,

$$
\begin{aligned}
\frac{1}{m} \sum_{k=1}^{m} E_x^{(\pi^1,\pi^2)} s(x_{k-1}, a_{k-1}, b_{k-1}) \geq{} & \frac{1}{\nu(W)} \frac{1}{m} \Big([E_x^{(\pi^1,\pi^2)} W(x_m) - W(x)] \\
& + (1-\lambda) E_x^{(\pi^1,\pi^2)} \sum_{k=1}^{m} W(x_{k-1}) \Big).
\end{aligned}
$$

The last inequality implies that

$$
\begin{aligned}
\liminf_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} E_x^{(\pi^1,\pi^2)} s(x_{k-1}, a_{k-1}, b_{k-1}) &\geq \frac{1-\lambda}{\nu(W)} \liminf_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} E_x^{(\pi^1,\pi^2)} W(x_{k-1}) \\
&\geq \frac{1-\lambda}{\nu(W)},
\end{aligned}
$$

proving the desired inequality. $\qquad\square$

The next corollary is a direct consequence of the two previous lemmas and Proposition 5.2.

Corollary 5.5 *Suppose Assumptions* 1 *and* 2 *hold. Then, for all* $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2, x \in \mathbf{X}$, *the following inequality holds*:

$$\liminf_{m \to \infty} \frac{1}{m} \sum_{k=0}^{m-1} E_x^{(\pi^1, \pi^2)} \tau(x_k, a_k, b_k) \geq \frac{(1-\lambda)l}{\nu(W)}.$$

Therefore,

$$\sum_{k=0}^{\infty} E_x^{(\pi^1, \pi^2)} \tau(x_k, a_k, b_k) = \infty$$

and

$$\lim_{m \to \infty} \frac{E_x^{(\pi^1, \pi^2)} u(x_m)}{\sum_{i=0}^{m-1} E_x^{(\pi^1, \pi^2)} \tau(x_i, a_i, b_i)} = 0$$

for all $u \in B_W(\mathbf{X})$.

Proof of Theorem 3.3 Fix an arbitrary strategy pair $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and an arbitrary initial state $x \in \mathbf{X}$. Assumption 1(b) and inequality (20) imply that

$$\limsup_{n \to \infty} \frac{1}{n} E_x^{(\pi^1, \pi^2)} \sum_{i=0}^{n-1} |C(x_i, a_i, b_i)| \leq \limsup_{n \to \infty} \frac{k}{n} E_x^{(\pi^1, \pi^2)} \sum_{i=0}^{n-1} W(x_i)$$

$$\leq \lim_{n \to \infty} \frac{k}{n} \left[\frac{1-\lambda^n}{1-\lambda} + \frac{n\nu(W)}{(1-\lambda)\nu(X)} \right] W(x)$$

$$= \frac{k}{l\nu(\mathbf{X})} \left(\frac{\nu(W)}{1-\lambda} \right)^2 W(x).$$

Then, from the above inequality and Corollary 5.5, it follows that

$$|J(\pi^1, \pi^2, x)| \leq \frac{\limsup_{n \to \infty} \frac{1}{n} E_x^{(\pi^1, \pi^2)} \sum_{i=0}^{n-1} |C(x_i, a_i, b_i)|}{\liminf_{n \to \infty} \frac{1}{n} E_x^{(\pi^1, \pi^2)} \sum_{i=0}^{n-1} \tau(x_i, a_i, b_i)}$$

$$\leq \frac{k(\nu(W))^2}{l(1-\lambda)^2 \nu(X)} W(x) < \infty, \tag{21}$$

which is the desired result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 5.6 *For all stationary strategy pair* $(\phi^1, \phi^2) \in \Phi^1 \times \Phi^2$:

$$|\rho(\phi^1, \phi^2)| \leq \frac{k}{l\nu(\mathbf{X})} \left(\frac{\nu(W)}{1-\lambda} \right)^2.$$

Hence, the constants

$$\rho^l := \sup_{\phi^2 \in \Phi^2} \inf_{\phi^1 \in \Phi^1} \rho(\phi^1, \phi^2) \quad and \quad \rho^u := \inf_{\phi^1 \in \Phi^1} \sup_{\phi^2 \in \Phi^2} \rho(\phi^1, \phi^2)$$

are well defined and finite; notice that $\rho^u \geq \rho^l$.

Proof It follows from (10), Proposition 3.1, Corollary 5.5 and Assumption 1(b) that

$$
\begin{aligned}
|\rho(\phi^1, \phi^2)| &\leq \frac{\mu_{\phi^1,\phi^2}(|C_{\phi^1,\phi^2}|)}{\mu_{\phi^1,\phi^2}(\tau_{\phi^1,\phi^2})} \\
&\leq \frac{k\nu(W)}{l(1-\lambda)}\mu_{\phi^1,\phi^2}(W) \\
&\leq \frac{k}{l\nu(\mathbf{X})}\left(\frac{\nu(W)}{1-\lambda}\right)^2,
\end{aligned}
$$

which is the result to be proven. $\qquad\square$

Now, to show the existence of solutions to the Shapley equation define the nonnegative kernel

$$\widetilde{Q}(B|x, a, b) := Q(B|x, a, b) - S(x, a, b)\nu(B)$$

for $(x, a, b) \in \mathbb{K}$, $B \in \mathcal{B}(\mathbf{X})$, and consider the operators on $B_W(\mathbf{X})$ given by

$$
\begin{aligned}
Lu(x, a, b) &:= C(x, a, b) - \rho^l \tau(x, a, b) + \widetilde{Q}u(x, a, b), \\
L^e u(x, a, b) &:= (Lu)^e(x, a, b), \\
Tu(x) &:= \sup_{\gamma^2 \in \mathbb{B}(x)} \inf_{\gamma^1 \in \mathbb{A}(x)} (L^e u)_{\gamma^1, \gamma^2}(x)
\end{aligned}
$$

for $(x, a, b) \in \mathbb{K}$, $B \in \mathcal{B}(\mathbf{X})$. Notice that $\widetilde{Q} \geq 0$ and also that

$$\sup_{(x,a,b)\in\mathbb{K}} \frac{|\widetilde{Q}u(x, a, b)|}{W(x)} \leq \lambda \|u\|_W \quad \forall u \in B_W(\mathbf{X}).$$

Lemma 5.7 *The operator T is a contraction operator from $L_W(\mathbf{X})$ into itself with modulus λ. Thus, by the Banach fixed-point theorem, there exists a unique function $h^* \in L_W(\mathbf{X})$ such that $h^* = Th^*$. Moreover*:

(a) $Th^*(x) = \inf_{\gamma^1 \in \mathbb{A}(x)} \sup_{\gamma^2 \in \mathbb{B}(x)} (L^e u)_{\gamma^1, \gamma^2}(x)$ *for all $x \in \mathbf{X}$;*
(b) *there exists $\phi^1_* \in \Phi^1$ such that*

$$h^*(x) = \sup_{\gamma^2 \in \mathbb{B}(x)} (L^e h^*)_{\phi^1_*, \gamma^2}(x) \quad \forall x \in \mathbf{X};$$

(c) *for each $\varepsilon > 0$ there exists $\phi^2_\varepsilon \in \Phi^2$ such that*

$$h^*(x) - \varepsilon \leq \inf_{\gamma^1 \in \mathbb{A}(x)} (L^e h^*)_{\gamma^1, \phi^2_\varepsilon}(x) \quad \forall x \in \mathbf{X}$$

Proof Let u be an arbitrary function in $L_W(\mathbf{X})$. It follows from Assumption 1 that $|Lu| \leq MW$, where $M := (k + |\rho^l|k + \lambda\|u\|_W)$. Then, since W is continuous, $|L^e u| \leq MW$, which in turn implies that $|(L^e u)_{\gamma^1, \gamma^2}| \leq MW$ for all measures $\gamma^1 \in \mathbb{A}(x)$, $\gamma^2 \in \mathbb{B}(x)$. For the other hand, $(L^e u)_{\gamma^1, \gamma^2}$ is a lower semicontinuous function on $\mathbf{X}$; hence, $(L^e u)_{\gamma^1, \gamma^2} \in L_W(\mathbf{X})$

for all measures $\gamma^1 \in \mathbb{A}(x)$, $\gamma^2 \in \mathbb{B}(x)$. Now, combining the above facts with the minimax theorem [10, Lemma 3.3], it follows that $Tu \in L_W(\mathbf{X})$ and also that

$$Tu(x) = \inf_{\gamma^1 \in \mathbb{A}(x)} \sup_{\gamma^2 \in \mathbb{B}(x)} (L^e u)_{\gamma^1, \gamma^2}(x) \ \forall x \in \mathbf{X}.$$

For the other hand, for $u, v \in B_W(\mathbf{X})$, notice that

$$||Tu - Tv||_W = \sup_{x \in \mathbf{X}} \frac{|Tu(x) - Tv(x)|}{W(x)}$$

$$\leq \sup_{x \in \mathbf{X}} \sup_{a \in A(x)} \sup_{b \in B(x)} \frac{|L^e u(x, a, b) - L^e v(x, a, b)|}{W(x)}$$

$$\leq \sup_{x \in \mathbf{X}} \sup_{a \in A(x)} \sup_{b \in B(x)} \frac{|Lu(x, a, b) - Lv(x, a, b)|}{W(x)}$$

$$\leq \sup_{x \in \mathbf{X}} \sup_{a \in A(x)} \sup_{b \in B(x)} \frac{|\widetilde{Q}(u - v)(x, a, b)|}{W(x)}$$

$$\leq \lambda ||u - v||_W.$$

Hence, T is a contraction operator from $L_W(\mathbf{X})$ into itself with modulus λ. So, there exists a unique function $h^* \in L_W(\mathbf{X})$ such that

$$h^*(x) = Th^*(x) = \inf_{\gamma^1 \in \mathbb{A}(x)} \sup_{\gamma^2 \in \mathbb{B}(x)} (L^e h^*)_{\gamma^1, \gamma^2}(x) \ \forall x \in \mathbf{X}.$$

The last two assertions follows from the minimax theorem [10, Lemma 3.3]. □

Lemma 5.8 $\nu(h^*) = 0$ *and* $\rho^l = \rho^u$.

Proof First it is shown that $\nu(h^*) \leq 0$ and that $\rho^l = \rho^u$; after this, it is proved that $\nu(h^*) \geq 0$, completing thus the proof.

For each $\varepsilon > 0$, by Lemma 5.7(b), there exists $\phi_\varepsilon^2 \in \Phi^2$ such that

$$h^* \leq C_{\phi^1, \phi_\varepsilon^2} - \rho^l \tau_{\phi^1, \phi_\varepsilon^2} + Q_{\phi^1, \phi_\varepsilon^2} h^* - \nu(h^*) S_{\phi^1, \phi_\varepsilon^2} + \varepsilon \ \ \forall \phi^1 \in \Phi^1.$$

Then, integrating with respect to the invariant probability measure $\mu_{\phi^1, \phi_\varepsilon^2}$ in both sides of the above inequality, it follows after a rearrangement of terms that

$$\rho^l \leq \rho(\phi^1, \phi_\varepsilon^2) + \frac{\varepsilon - \nu(h^*)\mu_{\phi^1, \phi_\varepsilon^2}(S_{\phi^1, \phi_\varepsilon^2})}{\mu_{\phi^1, \phi_\varepsilon^2}(\tau_{\phi^1, \phi_\varepsilon^2})} \ \ \forall \varepsilon > 0, \phi^1 \in \Phi^1.$$

Now suppose that $\nu(h^*) > 0$ and take $\varepsilon \in (0, \nu(h^*)\theta)$, where $\theta = (1 - \lambda)/\nu(W)$ (see Proposition 3.1(b)). Then

$$\rho^l \leq \rho(\phi^1, \phi_\varepsilon^2) + \frac{[\varepsilon - \nu(h^*)\theta]\nu(W)}{1 - \lambda} \ \ \forall \phi^1 \in \Phi^1,$$

which, taking infimum over ϕ^1, yields

$$\rho^l \leq \inf_{\phi^1 \in \Phi^1} \rho(\phi^1, \phi_\varepsilon^2) + \frac{[\varepsilon - \nu(h^*)\theta]\nu(W)}{1 - \lambda} < \rho^l.$$

Therefore, $\nu(h^*) \leq 0$.

Recall that C and S are lower semicontinuous and τ is continuous. Moreover, Qu is lower semicontinuous for each $u \in L_W(\mathbf{X})$. Now, since $\nu(h^*) \leq 0$, the function Lh^* is lower semicontinuous too. Hence,

$$L^e h^*(x, a, b) = Lh^*(x, a, b)$$

$$= C(x, a, b) - \rho^l \tau(x, a, b) + Qh^*(x, a, b) - \nu(h^*)S(x, a, b)$$

for all $(x, a, b) \in \mathbb{K}$. From Lemma 5.7(b), there exists $\phi_*^1 \in \Phi^1$ such that

$$h^*(x) = \sup_{\gamma^2 \in \mathbb{B}(x)} Lh^*_{\phi_*^1, \gamma^2}(x)$$

$$\geq C_{\phi_*^1, \phi^2}(x) - \rho^l \tau_{\phi_*^1, \phi^2}(x) + Q_{\phi_*^1, \phi^2} h^*(x) - \nu(h^*)S_{\phi_*^1, \phi^2}(x)$$

$$\geq C_{\phi_*^1, \phi^2}(x) - \rho^l \tau_{\phi_*^1, \phi^2}(x) + Q_{\phi_*^1, \phi^2} h^*(x) \qquad (22)$$

for all $x \in \mathbf{X}, \phi^2 \in \Phi^2$. Integrating with respect to $\mu_{\phi_*^1, \phi^2}$, it follows that

$$\rho^l \geq \rho(\phi_*^1, \phi^2) \quad \forall \phi^2 \in \Phi^2.$$

Thus,

$$\rho^l \geq \sup_{\phi^2 \in \Phi^2} \rho(\phi_*^1, \phi^2) \geq \inf_{\phi^1 \in \Phi^1} \sup_{\phi^2 \in \Phi^2} \rho(\phi^1, \phi^2) = \rho^u.$$

Therefore, $\rho^u = \rho^l$.

Next it is shown that $\nu(h^*) \geq 0$. Since $\rho^u = \rho^l$, the inequality (22) implies that

$$\frac{\nu(h^*)\mu_{\phi_*^1, \phi^2}(S_{\phi_*^1, \phi^2})}{\mu_{\phi_*^1, \phi^2}(\tau_{\phi_*^1, \phi^2})} \geq \rho(\phi_*^1, \phi^2) - \inf_{\phi^1 \in \Phi^1} \sup_{\phi^2 \in \Phi^2} \rho(\phi^1, \phi^2) \quad \forall \phi^2 \in \Phi^2;$$

thus,

$$\sup_{\phi^2 \in \Phi^2} \frac{\nu(h^*)\mu_{\phi_*^1, \phi^2}(S_{\phi_*^1, \phi^2})}{\mu_{\phi_*^1, \phi^2}(\tau_{\phi_*^1, \phi^2})} \geq \sup_{\phi^2 \in \Phi^2} \rho(\phi_*^1, \phi^2) - \inf_{\phi^1 \in \Phi^1} \sup_{\phi^2 \in \Phi^2} \rho(\phi^1, \phi^2) \geq 0.$$

This inequality and Proposition 5.1(c) imply that $\nu(h^*) \geq 0$. Therefore, $\nu(h^*) = 0$. $\qquad \square$

Proof of Theorem 3.4 Parts (a), (b) and (c) follow from Lemmas 5.7 and 5.8 with $\rho^* = \rho^l = \rho^u$. Now to prove part (d), observe that (b) implies that

$$h^*(x) \geq E^{(\phi_*^1, \pi^2)} \sum_{i=0}^{n-1} C(x_i, a_i, b_i) - \rho^* E^{(\phi_*^1, \pi^2)} \sum_{i=0}^{n-1} \tau(x_i, a_i, b_i) + E^{(\phi_*^1, \pi^2)} h^*(x_n)$$

for all $x \in \mathbf{X}, \pi^2 \in \Pi^2$. In turn, by Corollary 5.5, this inequality implies that

$$\rho^* \geq J(\phi_*^1, \pi^2, x) \quad \forall x \in \mathbf{X}, \pi^2 \in \Pi^2.$$

Hence,

$$\rho^* \geq \sup_{\pi^2 \in \Pi^2} J(\phi_*^1, \pi^2, x) \geq U(x) \quad \forall x \in \mathbf{X}.$$

Similarly, part (c) implies that

$$h^*(x) \leq E^{(\pi^1, \phi_\varepsilon^2)} \sum_{i=0}^{n-1} C(x_i, a_i, b_i) - \rho^* E^{(\pi^1, \phi_\varepsilon^2)} \sum_{i=0}^{n-1} \tau(x_i, a_i, b_i) + E^{(\pi^1, \phi_\varepsilon^2)} h^*(x_n) + n\varepsilon$$

for all $x \in \mathbf{X}, \pi^1 \in \Pi^1$, which, together with Corollary 5.5, implies that

$$\rho^* \leq J(\pi^1, \phi_\varepsilon^2, x) + \frac{\nu(W)l}{1 - \lambda} \varepsilon \quad \forall x \in \mathbf{X}, \pi^1 \in \Pi^1,$$

where l is the constant in Lemma 5.3. Thus,

$$\rho^* \leq \inf_{\pi^1 \in \Pi^1} J(\pi^1, \phi_\varepsilon^2, x) + \frac{\nu(W)l}{1 - \lambda} \varepsilon$$

$$\leq L(x) + \frac{\nu(W)l}{1 - \lambda} \varepsilon$$

for all $x \in \mathbf{X}$. Hence, since this inequality holds for all positive ε, it follows that

$$U(x) \leq \sup_{\pi^2 \in \Pi^2} J(\phi_*^1, \pi^2, x) \leq \rho^* \leq L(x) \quad \forall x \in \mathbf{X}.$$

Therefore, ρ^* is the value of the game and ϕ_*^1 is optimal for player 1.

Finally note that part (e) was already proved in Lemma 5.8. Thus, the proof of Theorem 3.4 is now complete. □

6 Proof of Theorem 4.2

The plan to prove Theorem 4.2 is to show that the inventory system satisfies all assumptions of Theorem 3.4, namely, Assumptions 1, 2 and 3. The next lemma is essential for the proof that Assumption 2 holds.

Lemma 6.1 *Suppose that Assumptions 4 and 5 hold. Then, there exists $r > 0$ such that (17) holds, that is, $\sup_{b \in \mathbf{B}} \Phi_b(r) < 1$ where $\Phi_p(\cdot)$ is as in (18).*

Proof A twice integration by parts leaves to the equality

$$e^{-x} = 1 - x + \frac{x^2}{2} - \frac{1}{2} \int_0^x (x - s)^2 e^{-s} ds \quad \forall x \geq 0;$$

this implies that

$$e^{-x} \leq 1 - x + \frac{x^2}{2} \quad \forall x \geq 0.$$

Let Z be a random variable with distribution $b \in \mathbf{B}$. Then,

$$\Phi_p(t) = E_b e^{t(\widehat{a} - Z)} = e^{t\widehat{a}} E_b e^{-tZ}$$

The last inequality implies that

$$\Phi_p(t) \le e^{t\widehat{a}} E_b (1 - t E_b Z + \frac{t^2}{2} E_b Z^2)$$

$$\le e^{t\widehat{a}} (1 - t z_* + \frac{t^2}{2} s^*) \quad \forall t \ge 0 \tag{23}$$

Now consider the function

$$h(t) := e^{-t\widehat{a}} - 1 + t z_* - \frac{t^2}{2} s^*, \quad t \in \mathbb{R},$$

and its derivative

$$h'(t) = -\widehat{a} e^{-t\widehat{a}} + z_* - t s^*.$$

Since h' is continuous and $h'(0) = z_* - \widehat{a} > 0$, there exists $\delta > 0$ such that $h' > 0$ on the interval $(0, \delta)$. For each $r \in (0, \delta)$, by the mean value theorem, there exists $t^* \in (0, r)$ such that

$$h(r) = h(r) - h(0) = h'(t^*) r > 0;$$

thus,

$$e^{-r\widehat{a}} - 1 + r z_* - \frac{r^2}{2} s^* > 0,$$

which implies that

$$\rho := e^{r\widehat{a}} (1 - r z_* + \frac{r^2}{2} s^*) < 1.$$

This latter fact combined with (23) yields that $\Phi_b(r) \le \rho$ for all $b \in \mathbf{B}$, which in turn implies the desired result. □

Lemma 6.2 *Suppose that Assumptions 4 and 5 hold. Let v be the Dirac measure at 0. Thus, the constant λ in Lemma 6.1, the measure v and the functions W and s given in (19) satisfy Assumption 2.*

Proof First observe Assumption 2(a) holds since $v(W) = W(0)$. For the other hand,

$$\int_{\mathbf{X}} v(y) Q(dy|x, a, b) = \int_{\mathbb{R}_+} v((x + a - w)^+) b(dw)$$

$$= \int_{[0, x+a]} v(x + a - w) b(dw) + v(0) b((x + a, \infty))$$

for all measurable function $v : \mathbf{X} \to \mathbb{R}$ whenever these integrals exist. In particular, it holds that

$$\int_{\mathbf{X}} W(y) Q(dy|x, a, b) = W(x) \int_{[0, x+a]} e^{r(a-w)} b(dw) + W(0) s(x, a, b)$$

$$\leq \Phi_b(r)W(x) + \nu(W)s(x,a,b)$$

$$\leq \lambda W(x) + \nu(W)s(x,a,b),$$

which proves that Assumption 2(c) holds. Similarly, taking $v = \mathbb{I}_B$, it results that

$$Q(B|x,a,b) = \int_{[0,x+a]} \mathbb{I}_B(x+a-w)b(dw) + \nu(B)b((x+a,\infty))$$

$$\geq \nu(B)s(x,a,b),$$

which is Assumption 2(b).

Finally, to prove the inequality $\nu(s_{\phi^1,\phi^2}) > 0$ for all stationary strategy pair (ϕ^1, ϕ^2), observe that

$$\nu(s_{\phi^1,\phi^2}) \geq \int_{\mathbf{B}} b((\widehat{a}, \infty))\phi^2(db|0).$$

Next, note that $b((\widehat{a}, \infty)) > 0$ because $\mu_b > \widehat{a}$ for all $b \in \mathbf{B}$, which implies that

$$\int_{\mathbf{B}} b((\widehat{a}, \infty))\phi^2(db|0) > 0 \quad \forall b \in \mathbf{B}.$$

Hence, $\nu(s_{\phi^1,\phi^2}) > 0$. $\square$

Lemma 6.3 *The payoff function C in* (13) *is lower semicontinuous.*

Proof Suppose that $(x_n, a_n, b_n) \to (x,a,b)$ and define

$$h_n(w) := (w - y_n)^+, \quad n \in \mathbb{N}, w \geq 0,$$

$$h(w) := (w - y)^+, \quad w \geq 0,$$

where $y_n := x_n + a_n, n \in \mathbb{N}$, and $y = x + a$; since $y_n \to y$, then $y_n \leq L$ for all $n \in \mathbb{N}$ for some $L > 0$. These functions satisfy the following properties:

(i) $\{h_n\}$ is asymptotically uniformly integrable with respect to the sequence of probability measures $\{b_n\}$, which means that

$$\lim_{K \to \infty} \limsup_{n \to \infty} \int_0^\infty |h_n(w)|\mathbb{I}_{B_n^K}(w)b_n(dw) = 0,$$

where $B_n^K := \{w \geq 0 : |h_n(w)| \geq K\}, n \in \mathbb{N}, K > 0$. In fact, note that

$$\int_{\mathbb{R}_+} |h_n(w)|\mathbb{I}_{B_n^K}(w)b_n(dw) = \int_{\mathbb{R}_+} (w - y_n)\mathbb{I}_{(y_n,\infty)}(w)\mathbb{I}_{(y_n+K,\infty)}(w)b_n(dw)$$

$$\leq \int_{\mathbb{R}_+} w\mathbb{I}_{[K,\infty)}(w)b_n(dw) - y_n b_n((L+K,\infty)).$$

Thus, taking limsup as n goes to ∞, it results that

$$\limsup_{n \to \infty} \int_{\mathbb{R}_+} |h_n(w)|\mathbb{I}_{B_n^K}(w)b_n(dw) \leq \int_{\mathbb{R}_+} w\mathbb{I}_{[K,\infty)}(w)b(dw) - yb((L+K,\infty)),$$

from which the desired result follows.

(ii) the sequence $\{h_n\}$ is equicontinuous; this property is verified directly.

(iii) $\{h_n\}$ converges to h in measure b; this follows because $h_n \to h$ pointwise.

Hence, by [1, Corollary 5.2], it holds that

$$E_b(w - x - a)^+ = \int_{\mathbb{R}_+} (w - x - a)^+ b(dw)$$

$$= \lim_{n \to \infty} \int_{\mathbb{R}_+} (w - x_n - a_n)^+ b_n(dw)$$

$$= \lim_{n \to \infty} E_{b_n}(w - x_n - a_n)^+.$$

Therefore,

$$\liminf_{n \to \infty} C(x_n, a_n, b_n) = \liminf_{n \to \infty} [c_1 \mathbb{I}_{(0,\infty)}(a_n) + c_2 a_n + c_3(x_n + a_n)$$

$$+ c_4 E_{b_n}(w - x_n - a_n)^+]$$

$$\geq C(x, a, b),$$

which proves that the cost function C is lower semicontinuous. □

Lemma 6.4 *W and QW are continuous functions.*

Proof Clearly, W is continuous. Now, to prove the second statement proceed as in the proof of the above lemma; thus, consider a sequence $(x_n, a_n, b_n) \in \mathbb{K}, n \in \mathbb{N}$, that converges to $(x, a, b) \in \mathbb{K}$. Next, define

$$\widehat{h}_n(\omega) := W((x_n + a_n - w)^+), \quad n \in \mathbb{N}, w \geq 0,$$

$$\widehat{h}(\omega) := W((x + a - w)^+), \quad w \geq 0.$$

The following facts hold:

(i) $\{\widehat{h}_n\}$ is asymptotically uniformly integrable with respect to the sequence of probability measures $\{b_n\}$, which means that

$$\lim_{K \to \infty} \limsup_{n \to \infty} \int_{\mathbb{R}_+} |\widehat{h}_n(w)| \mathbb{I}_{B_n^K}(w) b_n(dw) = 0,$$

where $B_n^K := \{w \geq 0 : |\widehat{h}_n(w)| \geq K\}$. This is true because the sequence $\{\widehat{h}_n\}$ is uniformly bounded.

(ii) the sequence $\{\widehat{h}_n\}$ is equicontinuous; to prove this, let L be a bound for the sequence $\{x_n + a_n\}$ and observe that the sequence of functions $g_n(w) := (x_n + a_n - w)^+, n \in \mathbb{N}, w \geq 0$, is equicontinuous. Since W is uniformly continuous in $[0, L]$, it follows that the sequence $\widehat{h}_n, n \in \mathbb{N}$, is equicontinuous too.

(iii) $\{\widehat{h}_n\}$ converges to $\widehat{h}$ in measure b; this follows from the pointwise convergence.

Hence, by [1, Corollary 5.2], it holds that

$$QW(x, a, b) = \int_{\mathbb{R}_+} \widehat{h}(\omega) b(dw)$$

$$= \lim_{n \to \infty} \int_{\mathbb{R}_+} \widehat{h}_n(\omega) b_n(dw)$$

$$= \lim_{n \to \infty} QW(x_n, a_n, b_n),$$

which proves the continuity of QW. $\square$

Lemma 6.5 *The transition law Q is weakly continuous.*

Proof Let $(x_n, a_n, b_n) \in \mathbb{K}, n \in \mathbb{N}$, be a sequence that converges to $(x, a, b) \in \mathbb{K}$; fix an arbitrary bounded continuous function $v : \mathbf{X} \to \mathbb{R}$ and an arbitrary real number $\varepsilon > 0$. The sequence $b_n, n \in \mathbb{N}$, is tight since it is weakly convergent; thus, there exists a constant $K_1 > 0$ such that

$$\sup_n b_n([K_1, \infty)) < \varepsilon/4M$$

where $M > 0$ is a bound for the function v. For the other hand, since $x_n + a_n \to x + a$, there exists $r > 0$ such that

$$0 \le x_n + a_n < K_2 := x + \widehat{a} + r \quad \forall n \in \mathbb{N}.$$

Now, let $K := \max\{K_1, K_2\}$ and observe that

$$(x_n + a_n - z)^+ = 0 \quad \forall z > K, n \in \mathbb{N}.$$

Moreover, since v is uniformly continuous on $[0, K]$, there exists $\delta > 0$ such that

$$|v(z_1) - v(z_2)| < \varepsilon/4$$

for all $z_1, z_2 \in [0, K]$ with $|z_1 - z_2| < \delta$.

Next consider the continuous bounded function

$$g(z) := v((x + a - z)^+), \quad z \in [0, \infty),$$

and take $N \in \mathbb{N}$ such that

$$|\int_0^\infty g(z) b_n(dz) - \int_0^\infty g(z) b(dz)| < \varepsilon/4,$$

$$|(x_n + a_n - z)^+ - (x + a - z)^+| < \delta \quad \forall z \in [0, \infty)$$

hold for all $n \ge N$, and put

$$A_n := |\int_\mathbf{X} v(y) Q(dy|x_n, a_n, b_n) - \int_\mathbf{X} v(y) Q(dy|x, a, b)|, \quad n \in \mathbb{N}.$$

Then,

$$A_n = |\int_0^\infty v((x_n + a_n - z)^+) b_n(dz) - \int_0^\infty v((x + a - z)^+) b(dz)|$$

$$\le |\int_0^\infty v((x_n + a_n - z)^+) b_n(dz) - \int_0^K v((x_n + a_n - z)^+) b_n(dz)| +$$

$$| \int_0^K v((x_n + a_n - z)^+)b_n(dz) - \int_0^K v((x + a - z)^+)b_n(dz)| +$$

$$\left| \int_K^\infty v((x + a - z)^+)|b_n(dz) \right| +$$

$$| \int_0^\infty v((x + a - z)^+)b_n(dz) - \int_0^\infty v((x + a - z)^+)b(dz)|.$$

Thus, for all $n \geq N$, it holds that

$$A_n \leq \int_K^\infty |v((x_n + a_n - z)^+)|b_n(dz) +$$

$$\int_0^K |v((x_n + a_n - z)^+) - v((x + a - z)^+)|b_n(dz) +$$

$$\int_K^\infty |v((x + a - z)^+)|b_n(dz) +$$

$$| \int_0^\infty v((x + a - z)^+)b_n(dz) - \int_0^\infty v((x + a - z)^+)b(dz)|,$$

which yields the inequality

$$A_n < 2Mb_n([K, \infty)) + \varepsilon G_n([0, K])/4 + \varepsilon/4 < \varepsilon \quad \forall n \geq N.$$

Hence, $Q(\cdot|x_n, a_n, b_n), n \in \mathbb{N}$, converges weakly to $Q(\cdot|x, a, b)$. $\qquad\square$

Lemma 6.6 *The set of probability distribution* **B** *is sequentially compact.*

Proof Notice that the following inequalities

$$z^* \geq \mu_b \geq \int_{(k, \infty)} sb(ds) \geq kb((k, \infty)) \tag{24}$$

hold for all $b \in \mathbf{B}$ and $k > 0$. Hence, **B** is tight.

Now take a sequence $\{b_n\} \subset \mathbf{B}$. By Prohorov theorem, there exist $b_0 \in \mathbb{P}(\mathbf{X})$ and a sub-sequence $\{b_{n_k}\}$ that converges weakly to b_0. Then,

$$z^* \geq \liminf_{k \to \infty} \int_0^\infty sb_{n_k}(ds) \geq \int_0^\infty sb_0(ds) = \mu_{b_0}$$

By (24), for each $\varepsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that

$$\sup_{b \in \mathbf{B}} \int_0^\infty s\mathbb{I}_{(k_0, \infty)}(s)b(ds) < \varepsilon.$$

Since the mapping $s \to s\mathbb{I}_{[0,k]}$ is upper semicontinuous and bounded from above, it follows for $m > k_0$ that

$$\mu_{b_0} \geq \int_0^\infty s\mathbb{I}_{[0,m]}(s)b_0(ds)$$

$$\geq \limsup_{n \to \infty} \int_0^\infty s\mathbb{I}_{[0,m]}(s)b_{n_k}(ds)$$

$$= \limsup_{n \to \infty} \left(\int_0^\infty s b_{n_k}(ds) - \int_0^\infty s \mathbb{I}_{(m,\infty)}(s) b_{n_k}(ds) \right)$$

$$\geq z_* - \varepsilon.$$

Because ε can be chosen arbitrarily, the last inequality implies that $\mu_{b_0} \geq z_*$.

Finally observe that

$$s^* \geq \liminf_{k \to \infty} \int_0^\infty s^2 b_{n_k}(ds) \geq \int_0^\infty s^2 b_0(ds).$$

Therefore, $b_0 \in \mathbf{B}$, which proves that $\mathbf{B}$ is sequentially compact. $\qquad \square$

Proof of Theorem 4.2 As mentioned at the beginning of this section, the plan is to show that the semi-Markov inventory system satisfies Assumptions 1, 2 and 3. Thus, first note that Remark 4.1 and Lemma 6.2 prove that Assumptions 1 and 2 hold, respectively. Moreover, Lemma 6.3 proves the lower semicontinuity of function C, which is Assumption 3(a). For the other hand, Assumption 3(d) follows from Assumption 6 since $\tau = \kappa$. Furthermore, Lemma 6.5 shows that Q is weakly continuous, which is Assumption 3(e), while Lemma 6.4 shows that Assumption 3(f) holds. Finally, note that Assumptions 3(b) and (c) trivially hold because $A(x) = [0, \widehat{a}]$ and $B(x) = \mathbf{B}$ for all $x \in \mathbf{X}$ and, according to Lemma 6.6, $\mathbf{B}$ is (sequentially) compact.

Next, consider a function $v : \mathbb{K} \to \mathbb{R}$ and observe that

$$\inf_{a \in A} \sup_{b \in B} v(x, a, b) = \inf_{v^1 \in \mathbb{A}} \sup_{v^2 \in \mathbb{B}} v(x, v^1, v^2) \quad \forall x \in \mathbf{X}.$$

Thus, from Theorem 3.4, there exists a function h^* in $L_W(\mathbf{X})$ such that

$$h^*(x) = \inf_{a \in A} \sup_{b \in B} [C(x, a, b) - \rho^* \tau(x, a, b) + Q h^*(x, a, b)] \quad \forall x \in \mathbf{X}.$$

Now, notice that for each $b \in \mathbf{B}$, the function

$$C(\cdot, \cdot, b) - \rho^* \tau(\cdot, \cdot, b) + Q h^*(\cdot, \cdot, b)$$

is lower semicontinuous. Then, the mapping

$$(x, a) \to \sup_{b \in B} [C(x, a, b) - \rho^* \tau(x, a, b) + Q h^*(x, a, b)]$$

is also lower semicontinuous. Hence, there exists $f_*^1 \in \mathbb{F}^1$ such that

$$h^*(x) = \sup_{b \in \mathbf{B}} [C(x, f_*^1(x), b) - \rho^* \tau(x, f_*^1(x), b) + \int_{\mathbf{X}} h^*(y) Q(dy | x, f_*^1(x), b)]$$

for all $x \in \mathbf{X}$, which proves the first statement of Theorem 4.2. The other statements are proved following standard arguments (as those given in the proof of Theorem 3.4(d) and (e)). $\qquad \square$

References

1. Feinberg, E.A., Kasyanov, P.O., Liang, Y.: Fatou's lemma in its classical form and Lebesgue's convergence theorems for varying measures with applications to Markov decision processes. Theory Probab. Appl. **65**, 270–291 (2020)
2. Gatsis, K., Ribeiro, A., Pappas, G.J.: Optimal power management in wireless control systems. IEEE Trans. Autom. Control **59**, 1495–1510 (2014)
3. González-Trejo, J.I., Hernández-Lerma, O., Hoyos-Reyes, L.F.: Minimax control of discrete-time stochastic systems. SIAM J. Control Optim. **41**, 1626–1659 (2002)
4. Guo, X.P., Zhu, Q.: Average optimality for Markov decision processes in Borel spaces: a new condition and approach. J. Appl. Probab. **43**, 318–334 (2006)
5. Hernández-Lerma, O., Lasserre, J.B.: Further Topics on Discrete-Time Markov Control Processes. Springer, New York (1999)
6. Hernández-Lerma, O., Vega-Amaya, O.: Infinite-horizon Markov control processes with undiscounted cost criteria: from average to overtaking optimality. Appl. Math. **25**, 153–178 (1998)
7. Hernández-Lerma, O., Vega-Amaya, O., Carrasco, G.: Sample-path optimality and variance-minimization of average cost Markov control processes. SIAM J. Control Optim. **38**(1), 79–93 (1999)
8. Jaśkiewicz, A.: Zero-sum semi-Markov games. SIAM J. Control Optim. **41**, 723–739 (2002)
9. Jaśkiewicz, A.: A fixed point approach to solve the average cost optimality equation for semi-Markov decision processes with Feller transition probabilities. Commun. Stat., Theory Methods **36**, 2559–2575 (2007)
10. Jaśkiewicz, A.: Zero-sum ergodic semi-Markov games with weakly continuous transition probabilities. J. Optim. Theory Appl. **141**, 321–347 (2009)
11. Jaśkiewicz, A., Nowak, A.S.: Zero-sum ergodic stochastic games with Feller transition probabilities. SIAM J. Control Optim. **45**, 773–789 (2006)
12. Jaśkiewicz, A., Nowak, A.S.: On the optimality equation for average cost Markov control processes with Feller transition probabilities. J. Math. Anal. Appl. **316**, 495–509 (2006)
13. Jaśkiewicz, A., Nowak, A.S.: Optimality in Feller semi-Markov control processes. Oper. Res. Lett. **34**, 713–718 (2006)
14. Jaśkiewicz, A., Nowak, A.S.: Robust Markov control processes. J. Math. Anal. Appl. **420**, 1337–1353 (2014)
15. Jaśkiewicz, A., Nowak, A.S.: Zero-sum stochastic games. In: Başar, T., Zaccour, G. (eds.) Handbook of Dynamic Game Theory. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-27335-8_8-1
16. Luque-Vásquez, F., Minjárez-Sosa, J.A., Rosas-Rosas, L.C.: Semi-Markov control models with partially known holding times distribution: discounted and average criteria. Acta Appl. Math. **114**, 135–156 (2011)
17. Mesquita, A.R., Hespanha, J.P., Nair, G.N.: Redundant data transmission in control/estimation over lossy networks. Automatica **48**, 1612–1620 (2012)
18. Nowak, A.S.: Measurable selection theorems for minimax stochastic optimization problems. SIAM J. Control Optim. **23**, 466–476 (1985)
19. Ross, S.M.: Applied Probability Models with Optimization Applications. Dover, New York (1970)
20. Tanaka, K., Wakuta, K.: On semi-Markov games. Sci. Rep. Niigata Univ. Ser. A **13**, 55–64 (1976)
21. Vega-Amaya, O.: The average cost optimality equation: a fixed point approach. Bol. Soc. Mat. Mex. **9**, 185–195 (2003)
22. Vega-Amaya, O.: Zero-sum semi-Markov games: fixed-point solutions of the Shapley equation. SIAM J. Control Optim. **42**, 1876–1894 (2003)
23. Vega-Amaya, O.: On the regularity property of semi-Markov processes with Borel state spaces. In: Hernández-Hernández, D., Minjárez-Sosa, A. (eds.) Optimization, Control, and Applications of Stochastic Systems, pp. 301–309. Springer, Berlin (2012)
24. Vega-Amaya, O.: Solutions of the average cost optimality equation for Markov decision processes with weakly continuous kernel: the fixed-point approach revisited. J. Math. Anal. Appl. **464**, 152–163 (2018)