Morfismos, Vol. 4, No. 2, 2000, pp. 15–31

Average optimal strategies in Markov games under a geometric drift condition *

Heinz-Uwe Küenle

Abstract

Zero-sum stochastic games with the expected average cost criterion and unbounded stage cost are studied. The state space is an arbitrary Borel set in a complete separable metric space but the action sets are finite. It is assumed that the transition probabilities of the Markov chains induced by stationary strategies satisfy a certain geometric drift condition. It is shown that the average optimality equation has a solution and that both players have optimal stationary strategies.

1991 Mathematics Subject Classification: 91A15
Keywords and phrases: Markov games, Borel state space, average cost criterion, geometric drift condition, unbounded costs

1 Introduction

In this paper two-person stochastic games with the expected average cost criterion are studied. The state space is a standard Borel space, that is, an arbitrary Borel set in a complete separable metric space. The action sets of both players are finite. Such a stochastic game can be described in the following way: The state x_n of a dynamic system is periodically observed at times $n = 1, 2, \ldots$. After an observation at time n the first player chooses an action a_n from the action set $\boldsymbol{A}(x_n)$ and afterwards the second player chooses an action b_n from the action set $\boldsymbol{B}(x_n)$ dependent on the complete history of the system at this time. The first player must pay cost $k^1(x_n, a_n, b_n)$, the second player must pay

*Invited article

$k^2(x_n, a_n, b_n)$, and the system moves to a new state x_{n+1} in the state space $\mathbf{X}$ according to the transition probability $p(\cdot \mid x_n, a_n, b_n)$.

Stochastic games with Borel state space and average cost criterion are considered by several authors. Related results are given by Maitra and Sudderth [7], [8], [9], Nowak [13], Rieder [15] and Küenle [6] in the case of bounded costs (payoffs). The case of unbounded payoffs is treated by Nowak [14] and Küenle [4]. The assumptions in this paper concerning the transition probabilities are related to Nowak's assumptions: Nowak assumes that there is a Borel set $C \in \mathbf{X}$ and for every stationary strategy pair $(\pi^\infty, \rho^\infty)$ a measure μ such that C is μ-small with respect to the Markov chain induced by this strategy pair. We assume that C is only a μ-petite set with respect to a resolvent of this Markov chain; as against this, we demand that μ is independent of the corresponding strategy pair. (For the definition of "small sets" and "petite sets" see [10].)

The paper is organised as follows: In Section 2 the mathematical model of Markov games is presented. Section 3 contains the assumptions on the transition probabilities and on the stage costs, and also some preliminary results. In Section 4 we study the expected average cost of a fixed stationary strategy pair. We show that the Poisson equation has a solution. In Section 5 we prove that the average cost optimality equation has a solution and both players have optimal stationary strategies.

2 The Mathematical Model

In this section we introduce the mathematical model of the stochastic game considered in this paper.

Definition 2.1
$\mathcal{M} = ((\mathbf{X}, \sigma_\mathbf{X}), (\mathbf{A}, \sigma_\mathbf{A}), \boldsymbol{A}, (\mathbf{B}, \sigma_\mathbf{B}), \boldsymbol{B}, p, k^1, k^2, \mathbf{E}, \mathbf{F})$ is called a *Markov game* if the elements of this tuple have the following meaning:

— $(\mathbf{X}, \sigma_\mathbf{X})$ is a standard Borel space, called the *state space.*

— $\mathbf{A}$ is a countable set and $\sigma_\mathbf{A}$ is the power set of $\mathbf{A}$. $\boldsymbol{A}(x) \in \mathbf{A}$ denotes a finite set of actions of the first player for every $x \in \mathbf{X}$. $\mathbf{A}$ is called the *action space of the first player* and $\boldsymbol{A}(x)$ is called the *admissible action set of the first player at state* $x \in \mathbf{X}$.

— $\mathbf{B}$ is a countable set and $\sigma_{\mathbf{B}}$ is the power set of $\mathbf{B}$. $\boldsymbol{B}(x) \in \mathbf{B}$ denotes a finite set of actions of the second player for every $x \in \mathbf{X}$. $\mathbf{B}$ is called the *action space of the second player* and $\boldsymbol{B}(x)$ is called the *admissible action set of the second player at state* $x \in \mathbf{X}$.

— p is a transition probability from $\sigma_{\mathbf{X}\times\mathbf{A}\times\mathbf{B}}$ to $\sigma_{\mathbf{X}}$, the *transition law.*

— k^i, $i = 1, 2$, are $\sigma_{\mathbf{X}\times\mathbf{A}\times\mathbf{B}}$-measurable functions, called *stage cost functions.*

— Let $\mathbf{H}_n = (\mathbf{X} \times \mathbf{A} \times \mathbf{B})^n \times \mathbf{X}$ for $n \geq 1$, $\mathbf{H}_0 = \mathbf{X}$. $h \in \mathbf{H}_n$ is called the *history at time n.*
A transition probability π_n from $\sigma_{\mathbf{H}_n}$ to $\sigma_{\mathbf{A}}$ with
$\pi_n(\boldsymbol{A}(x_n) \mid x_0, a_0, b_0, \dots, x_n) = 1$ for all $(x_0, a_0, b_0, \dots, x_n) \in \mathbf{H}_n$ is called a *decision rule of the first player at time n.*
A transition probability ρ_n from $\sigma_{\mathbf{H}_n\times\mathbf{A}}$ to $\sigma_{\mathbf{B}}$ with
$\rho_n(\boldsymbol{B}(x_n) \mid x_0, a_0, b_0, \dots, x_n) = 1$ for all $(x_0, a_0, b_0, \dots, x_n) \in \mathbf{H}_n$ is called a *decision rule of the second player at time n.*
A decision rule of the first [second] player is called *Markov* iff a transition probability $\tilde{\pi}_n$ from $\sigma_{\mathbf{H}_n}$ to $\sigma_{\mathbf{A}}$ [$\tilde{\rho}_n$ from $\sigma_{\mathbf{H}_n}$ to $\sigma_{\mathbf{B}}$] exists such that $\pi_n(\cdot \mid x_0, a_0, b_0, \dots, x_n) = \tilde{\pi}_n(\cdot \mid x_n)$ [$\rho_n(\cdot \mid x_0, a_0, b_0, \dots, x_n) = \tilde{\rho}_n(\cdot \mid x_n)$] for all $(x_0, a_0, b_0, \dots, x_n) \in \mathbf{H}_n \times \mathbf{A}$. (Notation: We identify π_n as $\tilde{\pi}_n$ and ρ_n as $\tilde{\rho}_n$.)
$\mathbf{E}$ and $\mathbf{F}$ denote non-empty sets of Markov decision rules.

A decision rule of the first [second] player is called *deterministic* if a function $e_n : \mathbf{H}_n \to \mathbf{A}$ [$f_n : \mathbf{H}_n \to \mathbf{B}$] exists such that $\pi_n(e_n(h_n) \mid h_n) = 1$ for all $h_n \in \mathbf{H}_n$ [$\rho_n(f_n(h_n) \mid h_n) = 1$ for all $(h_n) \in \mathbf{H}_n$].

A sequence $\Pi = (\pi_n)$ or $P = (\rho_n)$ of decision rules of the first or second player is called a *strategy* of that player.
Strategies are called *deterministic*, or *Markov* iff all their decision rules have the corresponding property.
A Markov strategy $\Pi = (\pi_n)$ or $P = (\rho_n)$ is called *stationary* iff $\pi_0 = \pi_1 = \pi_2 = \dots$ or $\rho_0 = \rho_1 = \rho_2 = \dots$. (Notation: $\Pi = \pi^\infty$ or $P = \rho^\infty$.) We assume in this paper that the sets of all admissible strategies are $\mathbf{E}^\infty$ and $\mathbf{F}^\infty$. Hence, only Markov strategies are allowed. But by means of dynamic programming methods it is also possible to get corresponding results for Markov games with larger sets of admissible strategies. If $\mathbf{E}$ and $\mathbf{F}$ are the sets of all Markov decision rules

(in the above sense) then we have a Markov game with perfect (or complete) information. In this case the action set of the second player may depend also on the present action of the first player. If $\mathbf{E}$ is the set of all Markov decision rules but $\mathbf{F}$ is the set of all Markov decision rules which do not depend on the present action of the first player then we have a usual Markov game with independent action choice. Let $\Omega := \mathbf{X} \times \mathbf{A} \times \mathbf{B} \times \mathbf{X} \times \mathbf{A} \times \mathbf{B} \times \dots$ and $K^{i,N}(\omega) := \sum_{j=0}^{N} k^i(x_j, a_j, b_j)$ for $\omega = (x_0, a_0, b_0, x_1, \dots) \in \Omega$, $i = 1, 2$, $N \in \mathbb{N}$. By means of the Ionescu-Tulcea Theorem (see, for instance, [11]), it follows that there exists a suitable σ-algebra $\mathcal{F}$ in Ω and for every initial state $x \in \mathbf{X}$ and strategy pair (Π, P), $\Pi = (\pi_n)$, $P = (\rho_n)$, a unique probability measure $\mathsf{P}_{x,\Pi,P}$ on $\mathcal{F}$ according to the transition probabilities π_n, ρ_n and p. Furthermore, $K^{i,N}$ is $\mathcal{F}$-measurable for all $i = 1, 2$, $N \in \mathbb{N}$. We set

$$V_{\Pi P}^{i,N}(x) = \int_{\Omega} K^{i,N}(\omega) \mathsf{P}_{x,\Pi,P}(d\omega) \tag{2.1}$$

and

$$\Phi_{\Pi P}^{i}(x) = \liminf_{N \to \infty} \frac{1}{N+1} V_{\Pi P}^{i,N}(x) \tag{2.2}$$

if the corresponding integrals exist.

Definition 2.2
A strategy pair (Π^*, P^*) is called a *Nash equilibrium pair* iff

$$\Phi^1_{\Pi^* P^*} \leq \Phi^1_{\Pi P^*}$$

$$\Phi^2_{\Pi^* P^*} \leq \Phi^2_{\Pi^* P}$$

for all strategy pairs (Π, P).

In this paper we will consider especially zero-sum Markov games, that means $k^1 = -k^2$. In this case we call a Nash equilibrium pair also an *optimal strategy pair*. We set $k := k^1$, $V_{\Pi P}^{N} := V_{\Pi P}^{1,N}$, $\Phi_{\Pi P} := \Phi^1_{\Pi P}$.

3 Assumptions and Preliminary Results

In this paper we use the same notation for a substochastic kernel and for the "expectation operator" with respect to this kernel, that means: If $(\mathbf{Y}, \sigma_{\mathbf{Y}})$ and $(\mathbf{Z}, \sigma_{\mathbf{Z}})$ are standard Borel spaces, $v : \mathbf{Y} \times \mathbf{Z} \to \mathbb{R}$ a

$\sigma_{\mathbf{Y}\times\mathbf{Z}}$-measurable function, and q a substochastic kernel from $(\mathbf{Y}, \sigma_{\mathbf{Y}})$ to $(\mathbf{Z}, \sigma_{\mathbf{Z}})$ then we put

$$qv(y) := \int_{\mathbf{Z}} q(dz \mid y) v(y, z) \text{ for all } y \in \mathbf{Y}$$

if this integral is well-defined.
Furthermore, we define the operator T by

$$Tu = k + pu$$

for all $\sigma_{\mathbf{X}}$-measurable $u : \mathbf{X} \to \mathbb{R}$ for which pu exists, that means,

$$Tu(x, a, b) = k(x, a, b) + \int_{\mathbf{X}} p(d\xi \mid x, a, b) u(\xi)$$

for all $x \in \mathbf{X}, a \in \mathbf{A}, b \in \mathbf{B}$.

Let $\Pi = (\pi_n) \in \mathbf{E}^\infty$, $P = (\rho_n) \in \mathbf{F}^\infty$. If $V^N_{\Pi P}$ exists, then we get

$$V^N_{\Pi P} = \pi_0 \rho_0 k + \sum_{j=1}^{N} \pi_0 \rho_0 p \cdots p \pi_j \rho_j k.$$

For $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$ we put $(\pi\rho p)^n := \pi\rho p(\pi\rho p)^{n-1}$ where $(\pi\rho p)^0$ denotes the identity. Let $\vartheta \in (0, 1)$. We set for every $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$, $x \in \mathbf{X}$, and $Y \in \sigma_{\mathbf{X}}$

$$Q_{\vartheta,\pi,\rho}(Y \mid x) := (1 - \vartheta) \sum_{n=0}^{\infty} \vartheta^n (\pi\rho p)^n \mathbf{I}_Y(x)$$

where $\mathbf{I}_Y$ is the characteristic function of the set Y.

We remark that for a stationary strategy pair $(\pi^\infty, \rho^\infty)$ the transition probability $Q_{\vartheta,\pi,\rho}$ is a resolvent of the corresponding Markov chain.

Assumption 3.1 There are: a nontrivial measure μ on $\sigma_{\mathbf{X}}$; a set $C \in \sigma_{\mathbf{X}}$; a $\sigma_{\mathbf{X}}$-measurable function $W \geq 1$; and constants $\vartheta \in (0, 1)$, $\alpha \in (0, 1)$, and $\beta \in \mathbb{R}$, with the following properties:

(a)

$$Q_{\vartheta,\pi,\rho} \geq \mathbf{I}_C \cdot \mu$$

for all $\pi \in \mathbf{E}$ and $\rho \in \mathbf{F}$,

(b)

$$pW \leq \alpha W + \mathbf{I}_C \beta,$$

(c)

$$\sup_{x \in \mathbf{X}, a \in \boldsymbol{A}(x), b \in \boldsymbol{B}(x)} \frac{|k(x,a,b)|}{W(x)} < \infty.$$

Assumption 3.1 (a) means that C is a "petite set", (b) is called "geometric drift towards C " (see Meyn and Tweedie [10]). We assume in this paper that Assumption 3.1 is satisfied.

Lemma 3.2 There are a $\sigma_{\mathbf{X}}$-measurable function V with $1 \leq W \leq V \leq W + const$, and a constant $\lambda \in (0,1)$ with

$$Q_{\vartheta,\pi,\rho} V \leq \lambda V + \mathbf{I}_C \cdot \mu V \tag{3.1}$$

and

$$\vartheta p V \leq \lambda V. \tag{3.2}$$

Proof: Without loss of generality we assume $\beta > 0$. Let $\beta' := \frac{\vartheta}{1-\vartheta}\beta$, $W' := W + \beta'$, and $\alpha' := \frac{\beta'+\alpha}{\beta'+1}$. Then it holds that $\alpha' \in (\alpha, 1)$ and

$$\begin{aligned} pW' &= pW + \beta' \\ &\leq \alpha W + \beta' + \beta \mathbf{I}_C \\ &\leq \alpha' W - (\alpha' - \alpha) W + \alpha' \beta' + (1 - \alpha')\beta' + \beta \mathbf{I}_C \\ &\leq \alpha' W' - (\alpha' - \alpha) + (1 - \alpha')\beta' + \beta \mathbf{I}_C \\ &= \alpha' W' + \beta' + \alpha - \alpha'(\beta' + 1) + \beta \mathbf{I}_C \\ &= \alpha' W' + \beta \mathbf{I}_C. \end{aligned} \tag{3.3}$$

Now let $W'' := W' - \beta' \mathbf{I}_C = W + \beta'(1 - \mathbf{I}_C)$. Then we get from (3.3)

$$\begin{aligned} p(W'' + \beta' \mathbf{I}_C) &= pW' \\ &\leq \alpha' W' + \beta \mathbf{I}_C \\ &= \alpha' W'' + \alpha' \beta' \mathbf{I}_C + \beta \mathbf{I}_C \\ &= \alpha' W'' + \alpha' \beta' \mathbf{I}_C + \frac{1-\vartheta}{\vartheta} \beta' \mathbf{I}_C \\ &= \alpha' W'' + \frac{\alpha' \vartheta + 1 - \vartheta}{\vartheta} \beta' \mathbf{I}_C \\ &\leq \alpha' W'' + \frac{\beta'}{\vartheta} \mathbf{I}_C. \end{aligned} \tag{3.4}$$

We put $\alpha'' := \frac{1-\vartheta}{1-\alpha'\vartheta}$. Then it holds that $\alpha' = \frac{\alpha''+\vartheta-1}{\alpha''\vartheta}$. For $\beta'' := \alpha''\beta'$ it follows:

$$pW'' \leq \frac{\alpha'' + \vartheta - 1}{\alpha''\vartheta} W'' - \frac{\beta''}{\alpha''} p\mathbf{I}_C + \frac{\beta''}{\alpha''\vartheta}\mathbf{I}_C.$$

Hence,

$$\alpha''\vartheta pW'' \leq (\alpha'' + \vartheta - 1)W'' - \vartheta\beta'' p\mathbf{I}_C + \beta''\mathbf{I}_C.$$

Then

$$(1-\vartheta)W'' \leq \alpha''W'' + \beta''\mathbf{I}_C - \vartheta p(\alpha''W'' + \ \beta''\mathbf{I}_C).$$

This implies

$$(1-\vartheta)W'' \leq \alpha''W'' + \beta''\mathbf{I}_C - \vartheta\pi\rho p(\alpha''W'' + \ \beta''\mathbf{I}_C)$$

for every $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$. Hence,

$$\begin{aligned} Q_{\vartheta,\pi,\rho}W'' &= \sum_{n=0}^{\infty}(1-\vartheta)\vartheta^n(\pi\rho p)^n W'' \\ &\leq \sum_{n=0}^{\infty}\vartheta^n(\pi\rho p)^n(\alpha''W'' + \beta''\mathbf{I}_C) \\ &\quad - \sum_{n=1}^{\infty}\vartheta^n(\pi\rho p)^n(\alpha''W'' + \beta''\mathbf{I}_C) \\ &= \alpha''W'' + \beta''\mathbf{I}_C. \end{aligned} \tag{3.5}$$

We choose $\vartheta' \in (\vartheta, 1)$ and set $\gamma := \max\{\frac{\beta''}{\mu(\mathbf{X})}, \frac{\beta'}{\vartheta'-\vartheta}\}$, $\lambda' := \frac{\alpha''+\gamma}{1+\gamma}$, $\lambda := \max\{\lambda', \vartheta'\}$. It follows that $\alpha'' < \lambda' \leq \lambda < 1$ and $\lambda' - \alpha'' = (1-\lambda')\gamma$. Hence,

$$(\lambda - \alpha'')W'' \geq \lambda' - \alpha'' \geq (1-\lambda')\gamma \geq (1-\lambda)\gamma. \tag{3.6}$$

We put $V := W'' + \gamma$. Obviously, $V \geq W'' \geq 1$ and $V \geq \gamma$. Then it follows

$$\begin{aligned} Q_{\vartheta,\pi,\rho}V &= Q_{\vartheta,\pi,\rho}W'' + \gamma \\ &\leq \alpha''W'' + \mathbf{I}_C \cdot \beta'' + \gamma \\ &\leq \alpha''W'' + \mathbf{I}_C \cdot \gamma\mu(\mathbf{X}) + \gamma \\ &\leq \alpha''W'' + \mathbf{I}_C \cdot \mu V + \gamma \\ &\leq \alpha''W'' + \mathbf{I}_C \cdot \mu V + (\lambda - \alpha'')W'' + \lambda\gamma \text{ (see (3.6))} \\ &= \lambda(W'' + \gamma) + \mathbf{I}_C \cdot \mu V \\ &= \lambda V + \mathbf{I}_C \cdot \mu V. \end{aligned}$$

Hence, (3.1) is proved.
From $\gamma \geq \frac{\beta'}{\vartheta' - \vartheta}$ it follows

$$\vartheta'\gamma \geq \vartheta\gamma + \beta'. \tag{3.7}$$

Then

$$\begin{aligned}
\vartheta p V &= \vartheta p W'' + \vartheta\gamma \\
&\leq \alpha'\vartheta W'' + \beta' + \vartheta\gamma \text{ (see (3.4))} \\
&\leq \alpha'\vartheta W'' + \vartheta'\gamma \text{ (see (3.7))} \\
&\leq \vartheta'(W'' + \gamma) \\
&= \vartheta' V \\
&\leq \lambda V.
\end{aligned}$$

Hence, (3.2) is also proved. $\square$

4 Properties of Stationary Strategy Pairs

For a function $u : \mathbf{X} \to \mathbb{R}$ we put $\|u\|_V := \sup_{x\in\mathbf{X}} \frac{|u(x)|}{V(x)}$. Furthermore, we denote by $\mathfrak{V}$ the set of all $\sigma_{\mathbf{X}}$-measurable functions u with $\|u\|_V < \infty$. In the following we will assume that on $\mathfrak{V}$ that metric is given which is induced by the weighted supremum norm $\|\cdot\|_V$. Then $\mathfrak{V}$ is complete.

Lemma 4.1 $\| \sup\limits_{n\in\mathbb{N},\pi\in\mathbf{E},\rho\in\mathbf{F}} (\pi\rho p)^n V\|_V < \infty.$

Proof: From Assumption 3.1(b) it follows that

$$(\pi\rho p)^n W \leq \alpha^n W + \frac{1}{1-\alpha}\beta.$$

By Lemma 3.2 we get

$$(\pi\rho p)^n V \leq (\pi\rho p)^n W + const \leq \alpha^n W + const' \leq \alpha^n V + const'.$$

The statement is implied by this. $\square$

Let T_w be the operator given by

$$T_w u(x,a,b) := (1-\vartheta)(\vartheta k(x,a,b) + w(x)) + \vartheta p u(x,a,b)$$

for all $u \in \mathfrak{V}$, $x \in \mathbf{X}$, $a \in \mathbf{A}$, $b \in \mathbf{B}$. We note that T_w has essentially the same structure as the cost operator T used in stochastic dynamic programming and stochastic game theory. This implies that some of our proofs are very similar to known proofs. Therefore we will restrict ourselves to a few remarks in these cases. (A very good exposition of basic ideas and recent developments in stochastic dynamic programming can be found in the books of Hernández- Lerma and Lasserre [1], [2].) Obviously,

$$T_w u = (1-\vartheta)\vartheta T(\frac{u}{1-\vartheta}) + (1-\vartheta)w. \tag{4.8}$$

Lemma 4.2 Let $w \in \mathfrak{V}$, $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$. Then the functional equation

$$u = \pi\rho T_w u \tag{4.9}$$

has a unique solution $u_w = S_{\pi\rho} w \in \mathfrak{V}$ and it holds:

$$S_{\pi\rho} w = \lim_{n\to\infty} (\pi\rho T_w)^n u = (1-\vartheta)\sum_{n=0}^{\infty} \vartheta^n (\pi\rho p)^n (\vartheta\pi\rho k + w) \tag{4.10}$$

for every $u \in \mathfrak{V}$.

Proof: We note that $\pi\rho T_w \mathfrak{V} \subseteq \mathfrak{V}$. From (3.2) it follows that $\pi\rho T_w$ is contracting on $\mathfrak{V}$ with modulus λ. The rest of the proof follows by Banach's Fixed Point Theorem. □

We define a new operator $S_{\gamma,\pi,\rho}$ by

$$S_{\gamma,\pi,\rho} w := -(1-\mathbf{I}_C)\gamma + S_{\pi\rho} w - \mathbf{I}_C \mu w \tag{4.11}$$

for $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$, $w \in \mathfrak{V}$ where $S_{\pi\rho}$ is the operator defined by the functional equation (4.9). The following lemma gives some properties of this operator.

Lemma 4.3 (a) $S_{\gamma,\pi,\rho}\mathfrak{V} \subseteq \mathfrak{V}$.

(b) $S_{\gamma,\pi,\rho}$ is isotonic.

(c) $S_{\gamma,\pi,\rho}$ is contracting.

Proof: (a) is obvious.

(b) Using (4.10) we get

$$\begin{aligned} S_{\gamma,\pi,\rho}w &= -(1-\mathbf{I}_C)\gamma + (1-\vartheta)\sum_{n=0}^{\infty}\vartheta^n(\pi\rho p)^n(\vartheta\pi\rho k + w) - \mathbf{I}_C\mu w \\ &= -(1-\mathbf{I}_C)\gamma + (1-\vartheta)\sum_{n=0}^{\infty}\vartheta^{n+1}(\pi\rho p)^n\pi\rho k \\ &\quad +(Q_{\vartheta,\pi,\rho} - \mathbf{I}_C\mu)w. \end{aligned} \tag{4.12}$$

The statement follows from Assumption 3.1 (a).

(c) By Lemma 3.2 and (4.12) we get for $u, v \in \mathfrak{V}$

$$\begin{aligned} |S_{\gamma,\pi,\rho}u - S_{\gamma,\pi,\rho}v| &= |(Q_{\vartheta,\pi,\rho} - \mathbf{I}_C\mu)(u-v)| \\ &\leq (Q_{\vartheta,\pi,\rho} - \mathbf{I}_C\mu)V\|u-v\|_V \\ &\leq \lambda V\|u-v\|_V. \ \square \end{aligned} \tag{4.13}$$

Lemma 4.4 The operator $S_{\gamma,\pi,\rho}$ has in $\mathfrak{V}$ a unique fixed point $u_{\gamma,\pi,\rho}$. $\mu u_{\gamma,\pi,\rho}$ is continuous and non-increasing in γ.

Proof: The existence and uniqueness of the fixed point follows from Lemma 4.3 by Banach's Fixed Point Theorem. From $S_{\gamma,\pi,\rho}v \geq S_{\gamma',\pi,\rho}v$ for $\gamma \leq \gamma'$, and the isotonicity of $S_{\gamma,\pi,\rho}$ it follows that $u_{\gamma,\pi,\rho} \geq u_{\gamma',\pi,\rho}$. Hence, $\mu u_{\gamma,\pi,\rho} \geq \mu u_{\gamma',\pi,\rho}$. Furthermore, for arbitrary γ, γ'

$$\begin{aligned} |u_{\gamma,\pi,\rho} - u_{\gamma',\pi,\rho}| &= |(1-\mathbf{I}_C)(\gamma'-\gamma) + (Q_{\vartheta,\pi,\rho} - \mathbf{I}_C\mu)(u_{\gamma,\pi,\rho} - u_{\gamma',\pi,\rho})| \\ &\leq |\gamma-\gamma'|V + \lambda\|u_{\gamma,\pi,\rho} - u_{\gamma',\pi,\rho}\|_V V \end{aligned}$$

Hence,

$$\|u_{\gamma,\pi,\rho} - u_{\gamma',\pi,\rho}\|_V \leq |\gamma-\gamma'| + \lambda\|u_{\gamma,\pi,\rho} - u_{\gamma',\pi,\rho}\|_V$$

and

$$\begin{aligned} |\mu u_{\gamma,\pi,\rho} - \mu u_{\gamma',\pi,\rho}| &\leq \|u_{\gamma,\pi,\rho} - u_{\gamma',\pi,\rho}\|_V \mu V \\ &\leq \frac{|\gamma-\gamma'|}{1-\lambda}\mu V. \quad \square \end{aligned}$$

Theorem 4.5 There exists a constant g and $v \in \mathfrak{V}$ such that

$$g + v = \pi\rho k + \pi\rho p v. \tag{4.14}$$

It holds:

$$g = \Phi_{\pi^\infty\rho^\infty}.$$

Proof: From Lemma 4.4 it follows that there is a γ^* with $\gamma^* = \mu u_{\gamma^*,\pi,\rho}$. Hence,

$$\begin{aligned} u_{\gamma^*,\pi,\rho} &= S_{\gamma^*,\pi,\rho} u_{\gamma^*,\pi,\rho} \\ &= -(1-\mathbf{I}_C)\gamma^* + S_{\pi\rho} u_{\gamma^*,\pi,\rho} - \mathbf{I}_C \mu u_{\gamma^*,\pi,\rho} \\ &= S_{\pi\rho} u_{\gamma^*,\pi,\rho} - \gamma^*. \end{aligned} \tag{4.15}$$

Let $w^* := u_{\gamma^*,\pi,\rho}$. If we put $w = w^*$ in (4.9), then we get

$$S_{\pi\rho} w^* = (1-\vartheta)(\vartheta\pi\rho k + w^*) + \vartheta\pi\rho p S_{\pi\rho} w^*.$$

It follows by (4.15) that

$$w^* + \gamma^* = (1-\vartheta)(\vartheta\pi\rho k + w^*) + \vartheta\pi\rho p(w^* + \gamma^*).$$

Therefore,

$$\vartheta w^* + (1-\vartheta)\gamma^* = (1-\vartheta)\vartheta\pi\rho k + \vartheta\pi\rho p w^*.$$

For $g = \frac{\gamma^*}{\vartheta}$, $v = \frac{w^*}{1-\vartheta}$ we get (4.14). From (4.14) it follows

$$Ng = \sum_{n=0}^{N-1} (\pi\rho p)^n \pi\rho k + (\pi\rho p)^N v - v.$$

If we consider Lemma 4.1 we get

$$g = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (\pi\rho p)^n \pi\rho k = \Phi_{\pi^\infty\rho^\infty}. \quad \square$$

5 Existence of optimal stationary strategies

We give first a lemma which concerns a certain auxiliary one-stage game. The results of this lemma are well-known and can be derived, for instance, from the results in [12].

Lemma 5.1 Let $u : \mathbf{X} \times \mathbf{A} \times \mathbf{B} \to \mathbb{R}$ a $\sigma_{\mathbf{X}\times\mathbf{A}\times\mathbf{B}}$-measurable function with $\sup_{x\in\mathbf{X},a\in\boldsymbol{A}(x),b\in\boldsymbol{B}(x)} \frac{|u(x,a,b)|}{V(x)} < \infty$. Then it holds:

(a) $\inf_{\pi\in\mathbf{E}} \sup_{\rho\in\mathbf{F}} \pi\rho u = \sup_{\rho\in\mathbf{F}} \inf_{\pi\in\mathbf{E}} \pi\rho u \in \mathfrak{V}$.

(b) There are $\pi^* \in \mathbf{E}$, $\rho^* \in \mathbf{F}$ with $\pi^*\rho u \le \pi^*\rho^* u \le \pi\rho^* u$ for all $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$.

For a function $v : \mathbf{X} \times \mathbf{A} \to \mathbb{R}$ ($v : \mathbf{X} \times \mathbf{B} \to \mathbb{R}$) we put $Lv := \inf_{\pi \in \mathbf{E}} \pi v$ ($Uv := \sup_{\rho \in \mathbf{F}} \rho v$). We can now prove the following lemma concerning an auxiliary functional equation.

Lemma 5.2 The functional equation

$$\begin{aligned} u &= \inf_{\pi \in \mathbf{E}} \sup_{\rho \in \mathbf{F}} \{(1-\vartheta)(\vartheta \pi \rho k + w) + \vartheta \pi \rho p u\} \\ &= LUT_w u \\ &= (1-\vartheta)\vartheta LUT(\frac{u}{1-\vartheta}) + (1-\vartheta) w \end{aligned} \tag{5.16}$$

has for every $w \in \mathfrak{V}$ a unique solution $u^* =: Sw$ in $\mathfrak{V}$.

Proof: Let $w \in \mathfrak{V}$. Then it follows from Lemma 5.1 that $LUT_w \mathfrak{V} \subseteq \mathfrak{V}$. Because $\pi \rho T_w$ is contracting on $\mathfrak{V}$, it holds for $u, v \in \mathfrak{V}$:

$$\pi \rho T_w u \leq \pi \rho T_w v + \lambda \|u - v\|_V V.$$

Since L and U are isotonic it follows:

$$LUT_w u \leq LUT_w v + \lambda \|u - v\|_V V.$$

Because u and v can be interchanged, we get that LUT_w is also contracting. The statement follows by Banach's Fixed Point Theorem. $\square$

In the following lemma $S_{\pi\rho}$ and S are the operators defined by the functional equations (4.9) and (5.16).

Lemma 5.3 For every $w \in \mathfrak{V}$ there are $\pi^* \in \mathbf{E}$, $\rho^* \in \mathbf{F}$ with

$$S_{\pi^*,\rho} w \leq Sw \leq S_{\pi,\rho^*} w \tag{5.17}$$

for all $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$. Furthermore,

$$Sw := \inf_{\pi \in \mathbf{E}} \sup_{\rho \in \mathbf{F}} S_{\pi\rho} w. \tag{5.18}$$

Proof: It follows from Lemma 5.1 that there are $\pi^* \in \mathbf{E}$, $\rho^* \in \mathbf{F}$ such that

$$\begin{aligned} \pi^* \rho T(\frac{u_w}{1-\vartheta}) &\leq LUT(\frac{u_w}{1-\vartheta}) \\ &\leq \pi \rho^* T(\frac{u_w}{1-\vartheta}) \end{aligned} \tag{5.19}$$

where $u_w = Sw$. Hence,

$$\pi^* \rho T_w u_w \leq LUT_w u_w = u_w \leq \pi \rho^* T_w u_w \tag{5.20}$$

for all $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$. Assume that

$$(\pi^* \rho T_w)^n u_w \leq u_w \leq (\pi \rho^* T_w)^n u_w \tag{5.21}$$

for $n \in \mathbb{N}$. Then it follows from (5.20) that

$$u_w \leq \pi \rho^* T_w((\pi \rho^* T_w)^n u_w = (\pi \rho^* T_w)^{n+1} u_w. \tag{5.22}$$

Analogously,

$$u_w \geq (\pi^* \rho T_w)^{n+1} u_w. \tag{5.23}$$

From (5.22) and (5.23) it follows by mathematical induction that (5.21) holds for all $n \in \mathbb{N}$. For $n \to \infty$ we get (5.17).
(5.18) follows immediately from (5.17). □

We define a new operator S_γ by

$$S_\gamma w := -(1 - \mathbf{I}_C)\gamma + Sw - \mathbf{I}_C \mu w$$

for $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$, $w \in \mathfrak{V}$, $\gamma \in \mathbb{R}$. The following lemma gives some properties of this operator.

Lemma 5.4 (a) $S_\gamma \mathfrak{V} \subseteq \mathfrak{V}$.

(b) S_γ is isotonic.

(c) S_γ is contracting with modulus λ.

(d) S_γ has in $\mathfrak{V}$ a unique fixed point v_γ. It holds $\lim_{n\to\infty}(S_\gamma)^n u = v_\gamma$ for every $u \in \mathfrak{V}$. Moreover, v_γ is isotonic and continuous in γ.

Proof: (a) is obvious.

(b) From (4.11) and (5.18) it follows that

$$S_\gamma w = \inf_{\pi \in \mathbf{E}} \sup_{\rho \in \mathbf{F}} S_{\gamma,\pi,\rho} w.$$

By Lemma 4.3 we get the statement.

(c) Let $w', w'' \in \mathfrak{V}$. By Lemma 5.3 it follows that there are $\pi'' \in \mathbf{E}$, $\rho' \in \mathbf{F}$, such that

$$Sw' \leq S_{\pi,\rho'} w'$$

$$Sw'' \geq S_{\pi'',\rho} w''$$

for all $\pi \in \mathbf{E}$, $\rho \in \mathbf{F}$. Hence,

$$\begin{aligned}
S_\gamma w' - S_\gamma w'' &= -(1-\mathbf{I}_C)\gamma + Sw' - \mathbf{I}_C \mu w' \\
&\quad -(-(1-\mathbf{I}_C)\gamma + Sw'' - \mathbf{I}_C \mu w'') \\
&\leq -(1-\mathbf{I}_C)\gamma + S_{\pi'',\rho'} w' - \mathbf{I}_C \mu w' \\
&\quad -(-(1-\mathbf{I}_C)\gamma + S_{\pi'',\rho'} w'' - \mathbf{I}_C \mu w'') \\
&= S_{\gamma,\pi'',\rho'} w' - S_{\gamma,\pi'',\rho'} w'' \\
&\leq \lambda V \|w' - w''\|_V
\end{aligned}$$

since $S_{\gamma,\pi'',\rho'}$ is contracting (see Lemma 4.3). Because w' and w'' can be interchanged, we get the statement.

(d) The existence of a unique fixed point $v_\gamma \in \mathfrak{V}$ and $\lim_{n\to\infty}(S_\gamma)^n u = v_\gamma$ for every $u \in \mathfrak{V}$ follows from Banach's Fixed Point Theorem. For $\gamma' \leq \gamma$ it holds

$$S_\gamma w \leq S_{\gamma'} w = S_\gamma w + (1-\mathbf{I}_C)(\gamma - \gamma') \leq S_\gamma w + (\gamma - \gamma')V.$$

Assume that for $n > 1$

$$S_\gamma^{n-1} v_{\gamma'} \leq v_{\gamma'} \leq S_\gamma^{n-1} v_{\gamma'} + \frac{\gamma - \gamma'}{1-\lambda} V.$$

Then it follows

$$S_\gamma^n v_{\gamma'} \leq S_{\gamma'} S_\gamma^{n-1} v_{\gamma'} \leq S_{\gamma'} v_{\gamma'} = v_{\gamma'} \leq S_{\gamma'}(S_\gamma^{n-1} v_{\gamma'} + \frac{\gamma - \gamma'}{1-\lambda} V)$$

$$\leq S_\gamma(S_\gamma^{n-1} v_{\gamma'} + \frac{\gamma - \gamma'}{1-\lambda} V) + (\gamma - \gamma')V$$

$$\leq S_\gamma^n v_{\gamma'} + \frac{\lambda(\gamma - \gamma')}{1-\lambda} V + (\gamma - \gamma')V \text{ (see (c))}$$

$$= S_\gamma^n v_{\gamma'} + \frac{\gamma - \gamma'}{1-\lambda} V.$$

Hence, by mathematical induction we find that this inequality holds for all $n \in \mathbb{N}$. For $n \to \infty$ it follows

$$v_\gamma \le v_{\gamma'} \le v_\gamma + \frac{\gamma - \gamma'}{1-\lambda} V.$$

The rest of the statement is implied by this. $\Box$

Theorem 5.5 There are $g = const$ and $v \in \mathfrak{V}$ with

$$g + v = LUTv. \tag{5.24}$$

It holds

$$g = \inf_{\Pi \in \mathbf{E}^\infty} \sup_{P \in \mathbf{F}^\infty} \Phi_{\Pi P}.$$

Furthermore, there is an optimal stationary strategy pair.

Proof: From Lemma 5.4 it follows that μv_γ is non-increasing in γ. Therefore, there is a γ^* with $\gamma^* = \mu v_{\gamma^*}$.

$$\begin{aligned} v_{\gamma^*} &= S_{\gamma^*} v_{\gamma^*} \\ &= -(1 - \mathbf{I}_C)\gamma^* + S v_{\gamma^*} - \mathbf{I}_C \mu v_{\gamma^*} \\ &= S v_{\gamma^*} - \gamma^*. \end{aligned} \tag{5.25}$$

Let $w^* := v_{\gamma^*}$. If we put $w = w^*$ in (5.16) then we get

$$Sw^* = LU((1-\vartheta)(\vartheta k + w^*) + \vartheta p S w^*).$$

It follows by (5.25)

$$w^* + \gamma^* = LU((1-\vartheta)(\vartheta k + w^*) + \vartheta p(w^* + \gamma^*)).$$

Therefore,

$$\vartheta w^* + (1-\vartheta)\gamma^* = LU((1-\vartheta)\vartheta k + \vartheta p w^*).$$

For $g = \frac{\gamma^*}{\vartheta}$, $v = \frac{w^*}{1-\vartheta}$ we get (5.24).

From (5.24) and Lemma 5.1 it follows that there are $\pi^* \in \mathbf{E}$, $\rho^* \in \mathbf{F}$, with

$$\pi^* \rho_n T v_{\gamma^*} - g \le v_{\gamma^*} \le \pi_n \rho^* T v_{\gamma^*} + \varepsilon - g$$

for all $\Pi = (\pi_n) \in \mathbf{E}^\infty$, $P = (\rho_n) \in \mathbf{F}^\infty$. It follows

$$\begin{array}{rcl} & & \pi^*\rho_0 T\pi^*\rho_1 T\cdots\pi^*\rho_N T v_{\gamma^*} - (N+1)g \\ \leq & v_{\gamma^*} & \leq \; \pi_0\rho^* T\pi_1\rho^* T\cdots\pi_N\rho^* T v_{\gamma^*} - (N+1)g \end{array}$$

For $N \to \infty$ we get

$$\Phi_{\Pi\rho^{*\infty}} \leq g \leq \Phi_{\pi^{*\infty}P}$$

for all $\Pi \in \mathbf{E}^\infty$, $P \in \mathbf{F}^\infty$. This implies

$$g = \inf_{\Pi\in\mathbf{E}^\infty} \sup_{P\in\mathbf{F}^\infty} \Phi_{\Pi P}$$

and the optimality of $(\pi^{*\infty}, \rho^{*\infty})$. □

Heinz-Uwe Küenle
Brandenburgische Technische Universität Cottbus
Institut für Mathematik
PF 101344
D-03013 Cottbus
GERMANY
Phone: +49 (0355) 69 3151
Fax: +49 (0355) 69 3164
kuenle@math.tu-cottbus.de

References

[1] Hernández-Lerma, O.; Lasserre, J. B.: *Discrete-Time Markov Control Processes: Basic Optimality Criteria.* Applications of Mathematics 30. Springer-Verlag, New York, 1996

[2] Hernández-Lerma, O.; Lasserre, J. B.: *Further Topics on Discrete-Time Markov Control Processes.* Applications of Mathematics 42. Springer-Verlag, New York, 1999

[3] Küenle, H.-U.: *Stochastische Spiele und Entscheidungsmodelle.* Teubner-Texte zur Mathematik 89. Teubner-Verlag, Leipzig, 1986

[4] Küenle, H.-U.: Stochastic games with complete information and average cost criterion. *Advances in Dynamic Games and Applications* [edited by J.A. Filar, V. Gaitsgory, K. Mizukami] (*Annals of the International Society of Dynamic Games,* Vol. 5) Birkhäuser, Boston, 2000, 325– 338

[5] Küenle, H.-U.: Equilibrium strategies in stochastic games with additive cost and transition structure. *International Game Theory Review,* 1, 1999, 131–147

[6] Küenle, H.-U.: On multichain Markov games. *Annals of the International Society of Dynamic Games.* Birkhäuser, to appear

[7] Maitra, A.; Sudderth, W.: Borel stochastic games with limsup payoff. *Ann. Probab.*, 21, 1993, 861–885

[8] Maitra, A.; Sudderth, W.: Finitely additive and measurable stochastic games. *Internat. J. Game Theory*, 22, 1993, 201–223

[9] Maitra, A.; Sudderth, W.: Finitely additive stochastic games with Borel measurable payoffs. *Internat. J. Game Theory*, 27, 1998, 257–267

[10] Meyn, S. P.; Tweedie, R. L.: *Markov Chains and Stochastic Stability.* Communication and Control Engineering Series. Springer-Verlag, London, 1993

[11] Neveu, J.: *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco 1965

[12] Nowak, A. S.: Minimax selection theorems. *J. Math. Anal. Appl.* 103, 1984, 106–116.

[13] Nowak, A. S.: Zero-sum average payoff stochastic games with general state space. *Games and Econ. Behavior* 7, 1994, 221–232

[14] Nowak, A. S.: Optimal strategies in a class of zero-sum ergodic stochastic games.*Math. Meth. Oper. Res.* 50, 1999, 399–419

[15] Rieder, U.: Average optimality in Markov games with general state space. *Proc. 3rd International Conf. on Approximation and Optimization,* Puebla, 1995