Stochastic games with metric state space

Citation for published version (APA):

Document status and date:
Published: 01/01/1978

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Stochastic games with metric state space

by

H.A.M. Couwenbergh

Memorandum COSOR 78-05

Eindhoven, February 1978

The Netherlands
Stochastic Games with Metric State Space

by

H.A.M. Couwenbergh

Abstract

In this paper the stochastic two person zero sum game of Shapley is considered, with metric state space and compact action spaces. It is proved that both players have stationary optimal strategies, under conditions which are weaker than those of Maitra and Parthasarathy (a.o. no compactness of the state space). This is done in the following way: we show the existence of optimal strategies first for the one-period game with general terminal reward, then for the n-period games \( n = 1, 2, \ldots \); further we prove that the game over the infinite horizon has a value \( v \), which is the limit of the n-period game values. Finally the stationary optimal strategies are found as optimal strategies in the one-period game with terminal reward \( v \).

1. Introduction

The stochastic games we consider are non-cooperative two person zero sum games with discrete time parameter (originally introduced by Shapley, ref. [9]). This means that a system is given with a set of states \( S \), and two so-called action spaces: \( A \) for player I, \( B \) for player II. The system is started at time \( t = 0 \) in a state \( s \in S \); both players choose an action: \( a \in A, b \in B \). As a consequence of these actions, player I receives a "reward" \( r(s,a,b) \) (this may be negative) from II, and the system moves to a new state \( s' \) according to a (sub-) probability measure \( p(\cdot | s,a,b) \) on \( S \) (in case \( p(\cdot | s,a,b) \) is defective, the game has a positive stopping probability). Then this process is repeated at time \( t = 1 \) from the new starting state \( s' \), and so forth. The reward at time \( t \) is multiplied with \( \beta^t \), where \( \beta > 0 \) is called the discount factor. Object of player I is to maximize the total expected discounted reward (over the infinite horizon), object of player II is to minimize this same amount. Sometimes these opposite ambitions can be met simultaneously, namely if there exists an optimal pair of strategies; then for each player it is not profitable to deviate from his optimal strategy. In this case the game
also has a "value" (in fact a real function on $S$, see section 2 for a
definition). In literature several sets of conditions have been given
for the existence of "stationary" optimal strategies. We mention results
of Vrieze [12] and Wessels [13] in case of a countable state space, and
of Maitra and Parthasarathy [5] and Parthasarathy [8] when $S$ is a com-
 pact metric space.

The theorems presented in this paper are obtained by combining and
generalizing the bounding function method of Wessels, and Maitra and
Parthasarathy's approach for a compact state space. The first theorem
(section 3) makes use of the continuity on $S$ of reward function $r$ and
transition law $p$, and produces a continuous value function; the second
one (section 4) uses measurability of $r$ and $p$ on $S$, causing the value
function to be measurable. In the proofs of both results we do not need
a contraction theorem in order to find the value. Section 5 contains an
extension and some remarks; we start in section 2, with a formal model
and preparations.

2. Model and prerequisites

First we have to give a definition of (sub-) transition probability. Let $X$
and $Y$ be metric spaces; denote by $\sigma_Y$ the Borel $\sigma$-algebra on $Y$. The map
$q : \sigma_Y \times X \to [0,1]$ (closed unity interval) is called a (sub-) transition
probability (abbreviation: (sub-) trpr) $X \to Y$ if $q(\cdot|x)$ is a measure on $Y$
with $q(Y|x) \leq 1$ for all $x \in X$, and $q(Y_0|\cdot)$ is a measurable function on $X$
for all $Y_0 \in \sigma_Y$ (by measurability we shall always mean Borel measurability).
If, moreover, $q(Y|x) = 1$, $x \in X$, then $q$ is called a (nondefective) trpr.
In this section we require that the state space $S$ and the action spaces $A$
and $B$ are nonempty metric spaces, that the transition law $p$ is a (sub-) trpr
$S \times A \times B \to S$, and that the reward $r$ is a measurable map $S \times A \times B \to \mathbb{R}$
(set of reals); $\beta$ is a positive number.

Define for $t \geq 1$ $H_t := S \times A \times B \times \cdots \times B$ ($t$ times $S \times A \times B$); an
element $h_t = (s_0, a_0, \ldots, b_{t-1})$ of $H_t$ is called history. Let $F$ be the set
of all trpr's $f$, $S \to A$, and $G$ the set of all trpr's $g$, $S \to B$; for $t \geq 1$ $\Pi_t(A)$
is the set of all trpr's $\pi_t$, $H_t \times S \to A$. Now $\Pi(A) := F \times \Pi_1(A) \times \Pi_2(A) \times \cdots$
defines the set of general strategies \( \pi \) for player I. A strategy \( \pi \) is called Markov if each \( \pi_t \) is independent of history, so \( \pi = (f_0, f_1, \ldots) \) with \( f_t \in F_0, t \geq 0 \). The set of Markov strategies is denoted by \( R(A) \).

If \( \pi = (f, f, \ldots) \) for some \( f \in F \), then \( \pi \) is said to be stationary; we write \( \pi = f^\infty \). Similarly for player II the set \( \Pi(B) \) of general strategies, the set \( R(B) \) of Markov strategies, and the stationary strategies \( g^\infty \) are defined. Take \( \Pi := \Pi(A) \times \Pi(B) \), \( R := R(A) \times R(B) \).

Let \( s_0 \in S \) and \( (\pi, \rho) \in \Pi \) be given. According to a theorem of Ionescu Tulcea (see Neveu [6], Prop. V 1.1; in order to apply this result \( \rho \) has to be nondefective, which may be accomplished by extending \( S \) with an extra state \(*\); take \( \rho(\ast|s,a,b) := 1 - \rho(s|s,a,b) \) and extend all real-valued functions \( h \) on \( S \) by \( h(\ast) := 0 \) this starting state and pair of strategies determine a (sub-) probability measure \( P_{s_0, \pi, \rho} \) with the following property. If \( E = s_0 \times a_0 \times \ldots \times b_0 \) is a measurable rectangle in \( H_{t+1} \) then

\[
P_{s_0, \pi, \rho}(E) = \int_{s_0} \int_{s_0} \int_{s_0} \int_{a_0} \int_{b_0} \cdots
\]

and this expression is measurable in \( s_0 \). This implies that \( P_{\pi, \rho} \) is a (sub-) trpr \( H_{t+1} \rightarrow S \) for all \( t \geq 0 \). We introduce the random variables \( X_t, K_t \) and \( L_t \) (\( t \geq 0 \)): these are respectively the state of the system, player I's action and player II's action, all at time \( t \).

Define for \( t \geq 0, s \in S \) and \( (\pi, \rho) \in \Pi \)

\[
\begin{align*}
E_s^{\pi, \rho} r^p_t(s, \pi, \rho) &:= E_s^{\pi, \rho} r^p_t(X_t, K_t, L_t) \\
E_s^{\pi, \rho} r^n_t(s, \pi, \rho) &:= E_s^{\pi, \rho} r^n_t(X_t, K_t, L_t) \quad (E_s^{\pi, \rho} \text{ denotes expectation w.r.t. } P_{s, \pi, \rho}; \text{ if } a \in \mathbb{R} \text{ then } a^+ = \max\{0, a\}, a^- = \max\{0, -a\}).
\end{align*}
\]

These expectations exist, but may equal infinity. From now on we assume that for all \( s, \pi \) and \( \rho \)

\[
\sum_{t=0}^{\infty} \beta^t r^p_t(s, \pi, \rho) < \infty \text{ or } \sum_{t=0}^{\infty} \beta^t r^n_t(s, \pi, \rho) < \infty, \quad \text{so } r^p_t(s, \pi, \rho) := E_s^{\pi, \rho} r^p_t(X_t, K_t, L_t) \quad \text{(expected reward at time } t),
\]

(continued...
\[ v_n(s, \pi, \rho) := \sum_{t=0}^{n-1} \beta^t r_t(s, \pi, \rho) \quad (n\text{-period total expected reward}, \, n = 1, 2, \ldots) \]

and

\[ v(s, \pi, \rho) := \lim_{n \to \infty} v_n(s, \pi, \rho) \quad \text{(total expected reward) are welldefined.} \]

Since \( P \) is a (sub-) trpr, \( r(s, \pi, \rho) \) is measurable in \( s \) (the integral of a quasi-integrable measurable function with a trpr is measurable in the remaining variables, see e.g. Neveu [6], Prop. III 2.1); so \( v_n(s, \pi, \rho) \) and \( v(s, \pi, \rho) \) are measurable too.

Note: in the remainder, if we omit the index for the state space in some formula, the function on \( S \) is meant; except in some obvious cases, all statements concerning such functions are supposed to hold pointwise.

Define for \( n \in \mathbb{N} \) (set of positive integers) and \( s \in S \)

\[
L_n(s) := \sup_{\pi \in \Pi(A)} \inf_{\rho \in \Pi(B)} v_n(s, \pi, \rho), \quad H_n(s) := \inf_{\pi \in \Pi(A)} \sup_{\rho \in \Pi(B)} v_n(s, \pi, \rho),
\]

and

\[
L(s) := \sup_{\pi \in \Pi(A)} \inf_{\rho \in \Pi(B)} v(s, \pi, \rho), \quad H(s) := \inf_{\pi \in \Pi(A)} \sup_{\rho \in \Pi(B)} v(s, \pi, \rho).
\]

It is easily checked that always \( L_n \leq v_n^H \leq v_n^L \leq v^H \); if \( L^L = v^H \), the game is said to have a value \( v \), where \( v^L \leq v \leq v^H \); if \( v^L = v^H \), the game is said to have a value \( v \), similarly for the \( n \)-period game values \( v_n \).

If there exists a pair \( (\pi^*, \rho^*) \in \Pi \) with \( v(\pi^*, \rho^*) \leq v(\pi^*, \rho^*) \leq v(\pi^*, \rho^*) \) for all \( \pi \) and \( \rho \) (this implies \( v^L = v^H \)), the strategies \( \pi^* \) and \( \rho^* \) are called optimal. When player I uses as criterion \( \phi(\pi) := \inf_{\rho} \mu(\pi, \rho) \) (minimal total expected reward) then evidently \( \pi^* \) maximizes \( \phi \), while the criterion \( \sup_{\rho} \mu(\pi, \rho) \) for II is minimized by \( \rho^* \). So in this respect \( \pi^* \) and \( \rho^* \) are really optimal.

In the following sections we shall prove that a pair of stationary optimal strategies exists; in order to achieve this, we need some more definitions and auxiliary results.

Let \( X \) be a metric space. If \( q \) is a measure on \( X \) and \( f \) is a quasi-integrable measurable map \( X \to \mathbb{R} \), we define \( qf := \int_X q(dx)f(x) \). By \( P_X \) the set of all probability measures on the Borel subsets of \( X \) is denoted. On \( P_X \) the so-called
weak topology may be constructed (see also below), which has the following properties (see e.g. Blackwell, Freedman and Orkin [1]).

**Lemma 2.1.** (Without proof). Let $X$ be a metric space.

i) (Definition) $q_n \to q$ in the weak topology on $P_X$ iff $q_n f \to q f$ for all bounded continuous $f : X \to \mathbb{R}$.

ii) Let $X$ be separable. Then $P_X$ is separable metric, the Borel $\sigma$-algebra on $P_X$ is the $\sigma$-algebra generated by the weak topology. This $\sigma$-algebra is also generated by the functions $q \mapsto q f$ for all bounded continuous $f$, or by all indicator functions $f$ of an arbitrary class of subsets of $X$ generating the Borel $\sigma$-algebra on $X$.

iii) If $X$ is a compact metric space, then $P_X$ is also compact metric.

If $Y$ is a separable metric space, then a trpr $q$, $X \to Y$, may be regarded as a Borel map $X \to P_Y$ (since $q(x) \in P_Y$, $x \in X$; Borel measurability from 2.1 ii)), and conversely.

We introduce a generalization of the weak topology on $P_S$.

Let $\mu$ be a measurable map $S \to (0,\infty)$. Define $V_\mu := \{u : S \to \mathbb{R} \mid |u(s)|/\mu(s) \text{ is bounded}\}$, $M_\mu := \{u \in V_\mu \mid u \text{ is measurable}\}$, $C_\mu := \{u \in M_\mu \mid u \text{ is continuous}\}$, and $\Pi_\mu := \{q \mid q \text{ is a measure on the Borel subsets of } S \text{ and } q \mu < \infty\}$. We define the *weak $\mu$-topology* on $\Pi_\mu$ as the smallest topology which makes the maps $q \mapsto q f$ ($q \in \Pi_\mu$) continuous for every $f \in C_\mu$. So $q_n \to q$ in the weak $\mu$-topology if and only if for all $f \in C_\mu$, $q_n f \to q f$ ($n \to \infty$).

The *strong $\mu$-topology* on $\Pi_\mu$ is defined as the smallest topology in which the maps $q \mapsto q f$ are continuous for all $f \in M_\mu$.

**Lemma 2.2.** Let $X$ and $Y$ be metric spaces. Assume $Y$ is compact, and $t$ is a continuous map $X \times Y \to \mathbb{R}$. Then $t$ is continuous in $x$ uniformly in $y$.

**Proof.** Straightforward.

**Lemma 2.3.** Let $X$ and $Y$ be metric spaces. Let $\Gamma$ be a compact valued function from $X$ to $Y$ (i.e. for all $x \Gamma(x)$ is a compact subset of $Y$). Suppose $\Gamma$ is continuous (that is: $\{x \mid \Gamma(x) \cap G \neq \emptyset\}$ and $\{x \mid \Gamma(x) \subset G'\}$ are open in $X$ whenever $G'$ is open) and $t : X \times Y \to \mathbb{R}$ is continuous. Then $\max_{y \in \Gamma(x)} t(x,y)$ is continuous in $x$. 

---

- 5 -
Proof. This lemma is Lemma 3.1 of Parthasarathy [8], see there.

3. Existence of optimal strategies, first case: continuity on S

In order to prove the existence of stationary optimal strategies, we proceed in the following way: first we prove that the one-period game with a terminal reward which is continuous and bounded with respect to the bounding function $\mu$ on $S$, has optimal one-step strategies; with this we construct optimal strategies for the $n$-period games; finally the stationary optimal strategies of the game over an infinite horizon are found, again by using our one-period game knowledge.

We state the main theorem first, and prove it via a series of lemmas.

Consider the game model determined by $S$, $A$, $B$, $p$, $r$, and $\beta$ and assume that a function $\mu$ with corresponding sets $V_\mu$, $M_\mu$, $C_\mu$ and $\Pi_\mu$ is given (see the previous section).

Theorem 3.1. Make the following conditions:

a) $S$, $A$ and $B$ are metric spaces, $A$ and $B$ compact;

b) $\mu$ is continuous on $S$;

c) $r$ is continuous on $S \times A \times B$;

d) $p$ is continuous on $S \times A \times B$ with respect to the weak $\mu$-topology on $\Pi_\mu$;

e) for $\overline{r}$, defined by $\overline{r}(s) := \inf_{\gamma \in \Pi_\mu} \sup_{\varphi \in \mathcal{P}_A} \int \int r(s,a,b) \varphi(da) \gamma(db)$, holds $\overline{r} \in V_\mu$;

f) there exists a $\beta' > 0$ such that for all $s$, $a$ and $b$, $\frac{1}{S} \int p(ds'|s,a,b) \mu(s') \leq \beta' \mu(s)$;

g) for all $(\pi, \rho) \in \Pi$ and $s \in S$, $\sum_{t=0}^{\infty} \beta^t r^P_t(s,\pi,\rho) < \infty$ or $\sum_{t=0}^{\infty} \beta^t r^n_t(s,\pi,\rho) < \infty$;

h) $\sum_{t=0}^{\infty} \beta^t |r_t(s,\pi,\rho)|/\mu(s) \to 0 \ (N \to \infty)$ uniformly on $S \times \Pi$.

Then the game has a value $v$ with $v = \lim_{n \to \infty} v_n$ and $v \in C_\mu$; moreover, there exist stationary optimal strategies.
From the conditions of this theorem it follows that $p$ is a (sub-) transition probability:

**Lemma 3.2.** Suppose $a$, $b$, $d$ and $f$ of theorem 3.1 hold, then $p$ is a (sub-) trpr $S \times A \times B \rightarrow S$.

**Proof.** Define for all $s$, $a$ and $b$ $q(s|a,b) \in \mathbb{P}_e$ ($e$ is the unity vector on $S$, $e(s) \equiv 1$) by

$$q(Z|s,a,b) := [\beta'\mu(s)]^{-1} \int_{Z} p(ds'|s,a,b)\mu(s') \quad (Z \text{ Borel in } S).$$

If $m$ is a nonnegative measurable function on $S$ then

$$\int_{S} \int_{S} p(ds'|s,a,b)m(s')\mu(s') = \beta'\mu(s) \int_{S} q(ds'|s,a,b)m(s'),$$

so we may express $p$ in $q$:

$$p(Z|s,a,b) = \beta'\mu(s) \int_{Z} q(ds'|s,a,b)/\mu(s').$$

This implies that if $q$ is a (sub-) trpr, then $p$ also.

Take $\Delta := \{K = 0 \cap C|0 \text{ is open in } S, C \text{ closed}\}$, $\Delta$ is a Boole semi-algebra which generates the Borel $\sigma$-algebra on $S$; in order to prove that $q$ is a (sub-) trpr, it suffices to prove that $q(K|s,a,b)$ is measurable on $S \times A \times B$ for every $K \in \Delta$ (see Neveu [6], p. 74).

Let $K = 0 \cap C$, $0^*$ and $C$ closed ($^*$ denotes the complement w.r.t. $S$). Let $d$ denote the metric distance on $S$; define $G_n := \{s \in S|d(s,C) < n^{-1}\}$ (here $d(s,C) = \inf\{d(s,s')|s' \in C\}$) and $H_n := \{s \in S|d(s,0^*) < n^{-1}\}$, $n \in \mathbb{N}$. According to K. Parthasarathy [7], Theorem 1.6, there exists for all $n$ a uniformly continuous map $f_n : S \rightarrow [0,1]$ such that $f_n(s) = 1$ if $s \in C$, 0 if $s \in G_n^*$; similarly we have $g_n : S \rightarrow [0,1]$ uniformly continuous with $g_n(s) = 1$ of $s \in 0^*$, 0 if $s \in H_n^*$. Let $g_n^* := e - g_n$, then $[f_n g_n^*\mu] \in C_\mu$, so
\[
\int_S p(ds' | s, a, b)f_n(s')g_n(s')\mu(s')
\]
is continuous on \(S \times A \times B\).

Since

\[
q(K | s, a, b) = \lim_{n \to \infty} \left[ \beta' \mu(s) \right]^{-1} \int_S p(ds' | s, a, b)f_n(s')g_n(s')\mu(s')
\]

\(q(K | s, a, b)\) is measurable on \(S \times A \times B\), q.e.d.

We use the following result on measurable selections.

**Lemma 3.3.** Let \(X\) and \(Y\) be metric spaces, \(Y\) compact; let for all \(x \in X\) \(D(x) \subset Y\) be nonempty such that \(K := \{(x,y) | x \in X, y \in D(x)\}\) is closed. Suppose \(w : K \to \mathbb{R}\) is continuous, then there exists a measurable map \(f : X \to Y\) with \(f(x) \in D(x)\) and \(w(x, f(x)) = \sup_{y \in D(x)} w(x, y)\) \((x \in X)\).

**Proof.** If \(X\) were also complete and separable, lemma 3.3 would be a simple application of Theorem 2.1 in Parthasarathy [8]. However, one may check that these two conditions are not necessary for the proof of Parthasarathy's theorem.

For convenience we introduce an operator notation. Let \(Q\) denote integration with respect to \(p(\cdot | s, a, b)\), and \(C(f, g)\) integration w.r.t. \(f(\cdot | s)\) and \(g(\cdot | s)\), if \(f \in F\) and \(g \in G\).

Define for all \(\pi \in \Pi(A)\) and \(h_1 \in H_1\) \(\pi[h_1] \in \Pi(A)\) by

\[
\pi[h_1] := (\pi_1(\cdot | h_1), \pi_2(\cdot | h_1), \ldots) \quad (\text{if } \pi = (f_0, f_1, \ldots) \in \Pi(A), \text{ this gives } \pi[h_1] = (f_1, f_2, \ldots)).
\]

Using a similar notation for player II we find by means of the Ionescu Tulcea integral evaluation (see section 2)

\[
r_{t+1}(\pi, \rho) = C(\pi_0, \rho_0)Q_{t}(\pi[\cdot], \rho[\cdot]) \quad (t \geq 0, (\pi, \rho) \in \Pi). \quad \text{From this follows easily that } \nu_{n+1}(\pi, \rho) = C(\pi_0, \rho_0)[r + 8Q_{n}(\pi[\cdot], \rho[\cdot])] \quad (n \in \mathbb{N}).
\]
First we consider the one-period game with certain terminal reward \( w \); if the players choose \( f \in F \) and \( g \in G \) then the total expected reward in this game is given by 
\[
L(f, g)w := C(f, g)[r + \beta Q w]
\]
(here \( w \) has to be a measurable map \( \mu \) to \( \overline{\mathbb{R}} \) (extended reals) which is bounded from below or above by some element of \( V \)). We define the "optimization" operator \( U \) by 
\[
(Uw)(s) := \inf_{g} \sup_{f} (L(f, g)w(s))(s \in S).
\]
The next result shows that \( Uw \) is indeed the value of this game if \( w \in C \); then also a pair of optimal one-step strategies \( (f^*, g^*) \) exists, with \( Uw = L(f^*, g^*)w \in C \).

**Lemma 3.4.** Let \( w \in C \). Under the assumptions a through f of theorem 3.1, there exist \( f^* \in F \) and \( g^* \in G \) such that 
\[
L(f^*, g^*)w \in C \quad \text{and for all } f \text{ and } g
\]
\[
L(f, g)w \leq L(f^*, g^*)w \leq L(f^*, g)w.
\]

**Proof.** Define 
\[
\mu_w(s, a, b) := r(s, a, b) + \beta \int_S p(ds' | s, a, b)w(s').
\]
On account of \( d \) and \( e \), \( \mu_w \) is continuous. Define on \( S \times P_A \times P_B \)
\[
K_w(s, \phi, \gamma) := \int_A \int_B \mu_w(s, a, b)\phi(da)\gamma(db),
\]
then for all \( f \in F \), \( g \in G \) we have 
\[
(L(f, g)w)(s) = K_w(s, f(s), g(s))\]
(here \( f \) and \( g \) are considered as Borel maps \( S \times P_A \) and \( S \times P_B \) respectively). \( K_w \) is continuous in \( (s, \phi, \gamma) \) (reasoning almost as in Maitra and Parthasarathy [5], Lemma 2.1; use our lemma 2.2). For fixed \( s \in S \) we may apply the minimax theorem of Sion ([10] Theorem 3.4) on \( K_w(s, *, *, *) \) and find
\[
\inf_{\phi} \sup_{\gamma} K_w(s, \phi, \gamma) = \inf_{\phi} \sup_{\gamma} K_w(s, \phi, \gamma).
\]
Define for \( s \in S \), \( \gamma \in P_B \)
\[
t(s, \gamma) := \max_{\phi} K_w(s, \phi, \gamma).
\]
According to lemma 2.3 \( t \) is continuous on \( S \times P_B \). By virtue of lemma 3.3 there exists a measurable map \( g^* : S \times P_B \) such that 
\[
t(s, g^*(s)) = \inf_{\gamma} t(s, \gamma) \quad (s \in S).
\]
So now we have 
\[
\inf_{\gamma} \sup_{\phi} K_w(s, \phi, \gamma) = \max_{\gamma} K_w(s, \phi, g^*(s)), \quad \text{both expressions continuous on } S
\]
(apply lemma 2.3 on \( t \)).
Similarly a Borel map \( f^* : S \times P_A \) is found, such that 
\[
\inf_{\phi} \sup_{\gamma} K_w(s, f^*(s), \gamma) = \min_{\gamma} K_w(s, f^*(s), \gamma).
\]
When we combine these results and state them in terms of $L(f, g)w$, we see that $L(f^*, g^*)w$ is continuous on $S$ and that $L(f, g^*)w \leq L(f^*, g^*)w \leq L(f^*, g)w$ ($f \in F$, $g \in G$). Finally $(L(f^*, g^*)w)(s) = \inf \sup K(U(s, \varphi, \gamma) \leq \tilde{r}(s) + \beta \beta'M'\mu(s)$ on account of conditions $e$ and $f$ ($M'$ is a bound for $|w(s)|/\mu(s)$); analogously $L(f^*, g^*)w \geq \tilde{r} - \beta \beta'M'\mu$, so $L(f^*, g^*)w \in V_\mu$.

Now we show the existence of optimal Markov strategies in the $n$-period games.

**Lemma 3.5.** For all $n \in \mathbb{N}$ there exist $n$-step Markov strategies $\pi^*_n$ and $\rho^*_n$ such that $v_n := v_n(\pi^*_n, \rho^*_n) \in C_\mu$ and $v_n(\pi, \rho^*_n) \leq v_n \leq v_n(\pi^*_n, \rho)$ for all $(\pi, \rho) \in \Pi$.

**Proof.** With induction. Let $n = 1$; we have $v_1(\pi, \rho) = L(\pi^*_0, \rho^*_0)0$ so the desired one-step optimal strategies are found by taking $w = 0$ in lemma 3.4. Suppose the statement to prove holds for some $n$. Applying lemma 3.4 with $w = v_n$ we find $f^*_n$ and $g^*_n$ such that $L(f^*_n, g^*_n)v_n$ has the properties mentioned there. Take $v_{n+1}(\pi^*_{n+1}, \rho) := f^*_n v_n(\pi^*_n, \rho)$ and let $\rho \in \Pi(B)$, then

\[
v_{n+1}(\pi^*_{n+1}, \rho) = C(f^*_n, \rho^*_0)[\tau + \beta Q(v_n(\pi^*_n, \rho^*_n))] \geq L(f^*_n, \rho^*_0)v_n \geq L(f^*_n, g^*_n)v_n = v_{n+1}(\pi^*_{n+1}, \rho^*_{n+1}).
\]

The rest of the proof is evident. □

From lemma 3.5 and the conditions $g$ and $h$ we deduce the existence of the game value $v$ and some properties of it.

**Theorem 3.6.** The game has a value $v$, $v = \lim_{n \to \infty} v_n$, $v \in C_\mu$ and $Uv = v$.

**Proof.** Let $\varepsilon > 0$. If $n$ is large enough, we have for all $s \in S$ $v^H(s) - v_n(s) = \inf_{\rho \in \Pi(B)} \sup_{\pi} \left[ v_n(s, \pi, \rho) + \sum_{t=n}^{\infty} \beta^t r_t(s, \pi, \rho) \right] - v_n(s) \leq \sup_{\pi} \sum_{t=n}^{\infty} \beta^t r_t(s, \pi, \rho^*_n(\rho)) \leq \varepsilon u(s)$, where $\rho$ is arbitrary in $\Pi(B)$. Analogously for the same $n$ $v^L - v_n \geq -\varepsilon u$. So $v := v^L = v^H = \lim_{n \to \infty} v_n$. Obviously $v \in C_\mu$ remains to prove $Uv = v$. Take $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $|v - v_n| \leq \varepsilon u$ if $n \geq N$ (modulus pointwise);
now \( Uv \leq \inf \sup C(f, g)(r + \beta Q(v_N^+ + \epsilon \mu)) \leq \inf \sup C(f, g)(r + \beta Q(v_N^+ + \epsilon \beta')) \)

since \( Uv_N = v_{N+1}^+ \). Similarly \( Uv \geq v - \epsilon \mu(1 + \beta') \), q.e.d. (note that \( U \) is not necessarily a contraction, in contrast with a similar operator used in Maitra and Parthasarathy [5]).

In order to prove the remaining part of theorem 3.1, we need the following lemma.

**Lemma 3.7.** Let \( \mu \) be a finite measure on the Borel subsets of a metric space \( X \). If \( t_n, n \in \mathbb{N} \), and \( t \) are measurable functions \( X \rightarrow \mathbb{R} \), if \( \mu|t| < \infty \) and \( t_n \leq t \) for all \( n \), then \( \mu[\limsup t_n] \geq \limsup \mu t_n \).

**Proof.** Directly from Fatou's lemma.

**Theorem 3.8.** There exist stationary optimal strategies.

**Proof.** The one-step strategies \( f^* \) and \( g^* \) which determine the desired optimal strategies, are found by applying lemma 3.4 with \( \mu = v \); we have \( v = Uv = L(f^*, g^*)v \) and \( L(f^*, g^*)v \leq v \leq L(f^*, g)v \) \( (f \in F, g \in G) \). Let \( K \in \mathbb{N} \) be such that \( |v - v_n| \leq \mu \) if \( n \geq K \) (see above); for all \( n \in \mathbb{N} \), define \( \pi^N := (\pi^0, \ldots, \pi^N_{n-1}) \) \((n \in \mathbb{N})\).

First we prove with induction for all \( n \in \mathbb{N} \) the statement \( B_N \):

\[
\forall \pi \in \Pi(A) \left[ w_\pi(n, N) := v_{n+N}(\pi^N_{\pi(n)}, g^* \ldots g^* \phi(n)) \leq v + (\beta')^N \mu \right.
\]

if \( n \geq K \), and \( \limsup w_\pi(n, N) \leq v \).

For \( N = 1 \): let \( \pi \in \Pi(A) \), then \( w_\pi(n, 1) = L(\pi_0, g^*)v_n \leq C(\pi_0, g^*)(r + \beta Q(v + \mu)) \leq v + \beta \mu \) \((n \geq K)\), and using lemma 3.7, \( \limsup w_\pi(n, 1) = L(\pi_0, g^*) \limsup v_n \leq v \).

Assume that \( B_N \) holds for some \( N \). We have for all \( \pi \in \Pi(A) \) \( w_\pi(n, N + 1) = \)

\[
L(\pi_0, g^*)v_{n+N}(\pi^N_{\pi(n)}, g^* \ldots g^\phi(n)) \leq v + (\beta')^{N+1} \mu \quad (n \geq K),
\]

and by the second part of the induction assumption, \( \limsup w_\pi(n, N + 1) \leq v \).
So $B_{N+1}$ has been proved.

Let now $\pi \in \Pi(A)$ and $\varepsilon > 0$. It follows from $B_N$ that

\[ v \geq \lim_{n \to \infty} \sup \left[ v_N(\pi, g^{*\infty}) + \sum_{t=1}^{n+N} r_t(\pi, \pi_t(n), g^{*t}, g^{*t}, \ldots, g^{*t}, \ldots) \right] \geq v(\pi, g^{*\infty}) - 2\varepsilon \mu , \]

if $N$ is large enough, on account of condition $h$. So we may conclude that $v \geq v(\pi, g^{*\infty})$. Analogously $v \leq v(f^{*\infty}, \rho)$ for all $\rho \in \Pi(B)$ can be proved.

It follows that $(f^{*\infty}, g^{*\infty})$ is optimal.

4. Existence of optimal strategies, second case: measurability on $S$

This section provides an analogue of the previous section for the case in which all relevant functions on $S$ are measurable, like they were continuous there.

Again we assume that $S$, $A$, $B$, $p$, $r$, $\beta$, $\mu$, $V_\mu$, $M_\mu$, $C_\mu$, and $\Pi_\mu$ as in section 2 are given. We impose a stronger condition on $S$, namely that $S$ is a nonempty Borel subset of a Polish (that is: complete separable metric) space (also named standard Borel space, abbr. SB-space): this is necessary in view of the different selection theorem we use here.

**Theorem 4.1.** Suppose $S$ is an SB-space, $A$ and $B$ are compact metric; $\mu$ is measurable; $r(s,a,b)$ is measurable in $s$ for fixed $(a,b)$ and continuous in $(a,b)$ for fixed $s$; $p$ is a (sub-) trpr $S \times A \times B \to S$, and for fixed $s$ $p(\cdot | s, a, b)$ is continuous in $(a,b)$ w.r.t. the strong $\mu$-topology on $\Pi_\mu$; $\varepsilon$, $f$, $g$ and $h$ as in theorem 3.1.

Then the game possesses a value $v$, $v = \lim_{n \to \infty} v_n$, and $v \in M_\mu$; there exist stationary optimal strategies.

Note: here we could not deduce from the other conditions that $p$ is a (sub-) trpr, as we did in lemma 3.2.

The next result is useful for the proof of theorem 4.1.

**Lemma 4.2.** Let $X$ and $Y$ be metric spaces, $Y$ separable; suppose that the map $f : X \times Y \to \mathbb{R}$ has the following properties: $f(\cdot, y)$ is measurable on $X$ for all $y \in Y$, and $f(x, \cdot)$ is continuous on $Y$ for all $x \in X$. Then $f$ is measurable on $X \times Y$. 

Proof. Let \( \{a_n | n \in \mathbb{N}\} \) be a countable subset which lies dense in \( Y \).

For each \( y \in Y \) we may construct a subsequence \( \{b_n(y)\} \) of \( \{a_n\} \) such that 
\[
b_n(y) \to y \quad (n \to \infty)
\]
for all \( y \), and \( b_n(y) \) is a Borel measurable function of \( y \) \((n \in \mathbb{N})\). Define \( f_n(x,y) := f(x,b_n(y)) \). Now \( f = \lim_{n \to \infty} f_n \) (pointwise),
and \( f_n \) is measurable on \( X \times Y \). Consequently, \( f \) is measurable also.

As in the previous section, we need a selection theorem.

**Lemma 4.3.** Let \( X \) be an SB-space and \( Y \) a compact metric space. Let for all \( x \in X \) \( D(x) \subset Y \) be nonempty such that \( K := \{(x,y) | x \in X, y \in D(x)\} \) is closed in \( X \times Y \). If \( w \) is a measurable map \( K \to \mathbb{R} \) and \( w(x,y) \) is continuous in \( y \) for fixed \( x \in X \), then there exists a measurable function \( f : X \to Y \) with \( f(x) \in D(x) \) and \( w(x,f(x)) = \sup_{y \in D(x)} w(x,y) \) \((x \in X)\).

**Proof.** Application of Corollary 1 of Brown and Purves [2].

Note that lemma 4.3 is not a generalization of lemma 3.3 since there \( X \) is only metric. Using the same operator notation as in section 3, we proceed with the analogue of lemma 3.4.

**Lemma 4.4.** If \( w \in M_\mu \) and if the conditions \( a \) through \( f \) of theorem 4.1 are fulfilled, then there exist \( f^* \in F \) and \( g^* \in G \) with \( L(f^*,g^*) w \in M_\mu \) and for all \( f \) and \( g \) \( L(f,g^*) w \leq L(f^*,g^*) w \leq L(f^*,g) w \).

**Proof.** By lemma 4.2, \( \ell_w \), as defined in lemma 3.4, is measurable on \( S \times A \times B \) and continuous on \( A \times B \) for fixed \( s \in S \). It follows that
\[
K_w(s,\varphi,\gamma) = \int_A \int_B \ell_w(s,a,b) \varphi(da) \gamma(db)
\]
is measurable in \( s \) for all \( \varphi \in P_A \) and \( \gamma \in P_B \), and continuous in \( (\varphi,\gamma) \) for fixed \( s \) (proof of the latter as in Maitra and Parthasarathy [5], Lemma 2.1).

Applying Sion's minimax theorem, we get for all \( s \in S \)
\[
\sup \inf \ K_w(s,\varphi,\gamma) = \inf \sup \ K_w(s,\varphi,\gamma).
\]
Define on \( S \times P_B \) \( t(s,\gamma) := \max_{\varphi} K_w(s,\varphi,\gamma) \). Fix \( \gamma \in P_B \). We apply lemma 4.3
on $T(s,\varphi) := K_w(s,\varphi,\gamma)$ and find a measurable $f^0 : S \rightarrow P_A$ with $T(s, f^0(s)) = t(s, \gamma)$, $s \in S$; $T$ is measurable in $(s,\varphi)$ and $(s, f^0(s))$ is a measurable function of $s$, so $T(s, f^0(s))$ is measurable. Hence, $t(s, \gamma)$ is measurable in $s$ for fixed $\gamma$.

Using lemma 2.3 with $s$ fixed, we see that $t(s, \cdot)$ is continuous on $P_B$.

Lemma 4.3 provides now a measurable function $g^* : S \rightarrow P_B$ satisfying
\[ t(s, g^*(s)) = \inf_{\gamma} t(s, \gamma) \quad (s \in S). \]
Since $t$ is measurable on $S \times P_B$, $t(s, g^*(s))$ is measurable in $s$. So $\inf_{\gamma} \sup_{\varphi} K_w(s, \varphi, \gamma) = \max_{\gamma} K_w(s, \varphi, g^*(s))$,
both expressions are measurable in $s$. Analogously an $f^* \in F$ is found with
\[ \sup_{\varphi} \inf_{\gamma} K_w(s, \varphi, \gamma) = \min_{\gamma} K_w(s, f^*(s), \gamma). \]
If we combine these results and remember that $(L(f, g)w)(s) = K_w(s, f(s), g(s))$ for all $s$, $f$ and $g$, we see that
\[ L(f^*, g^*)w \leq L(f^*, g)w \leq L(f^*, g)w \quad (f \in F, g \in G). \]
The proof of $L(f^*, g^*)w \in V_\mu$ is the same as in lemma 3.4.

The proof of theorem 4.1 proceeds in a perfectly similar way as the proof of theorem 3.1; we only have to replace $C$ by $M$ in the text of section 3, starting with lemma 3.5.

Note that the conditions of theorem 3.1 are not stronger than those of theorem 4.1: in this section $S$ is an SB-space, and both conditions are incomparable.

5. Extension and other remarks

The results of the foregoing sections may be extended in the following direction (as is done in Parthasarathy [8]). Let for all $s \in S$ be given a nonempty Borel subset $A(s)$ of $A$ (the set of "admissible" actions) and a nonempty Borel subset $B(s)$ of $B$. We denote by $\tilde{\Pi}(A)$ the set of strategies $\pi$ which restrict themselves to admissible actions, so $\pi_t(A(s)|h_t, s) = 1$ for all $t \geq 0$, $h_t \in H_t$ and $s \in S$. Analogously $\tilde{\Pi}(B)$ is defined; $\tilde{\Pi} := \tilde{\Pi}(A) \times \tilde{\Pi}(B)$.

We can extend theorem 3.1 to the game with admissible actions:

**Theorem 5.1.** Replace in theorem 3.1 $P_A$ by $P_A(s)$, $P_B$ by $P_B(s)$, $g$ and $h$ by $\tilde{\Pi}$. Make the following additional assumptions: $A(s)$ and $B(s)$ are compact, $s \in S$; the set valued functions $Q_A(s) := P_A(s)$ and $Q_B(s) := P_B(s)$
are continuous (that is: \( \{ s | Q_A(s) \cap G' \neq \emptyset \} \) and \( \{ s | Q_A(s) \subseteq G' \} \) are open in \( S \) if \( G' \) is open in \( P_A \), similarly for \( Q_B \)). Then, within the class \( \tilde{\Pi} \), the game possesses a value \( v \) with \( v = \lim_{n \to \infty} v_n \) and \( v \in \mathbb{C}_v \), and there exist stationary optimal strategies.

Proof. If lemma 3.4 produces an \( f^* \) and a \( g^* \) which satisfy \( f^*(s) \in P_A(s)' \) and \( g^*(s) \in P_B(s)' \), \( s \in S \), then we can restrict lemma 3.5 and theorem 3.6 and 3.8 to \( \tilde{\Pi} \). And this is not difficult to check, since the lemmas 2.3 and 3.3 are already stated in the proper form for this extension (when applying lemma 3.3, \( K \) is closed because \( Q_A \) and \( Q_B \) are continuous, so upper semi-continuous; see Parthasarathy [8]).

Similarly of course, theorem 4.1 may be extended.

We conclude with three remarks.

Remark 5.2. A slightly different version of theorem 3.1 can be proved, where \( S \) is supposed to be an \( SB \)-space, but in condition \( h \) \( \Pi \) is replaced by \( R \). We use the following theorem, which is based on an idea of Strauch [11]. Let \( (\pi, \rho) \in \Pi \); if one of these strategies, say \( \rho \), is Markov, then for each starting state \( s \in S \) there exists a Markov strategy for the other player, say \( \pi^* \), such that the (sub-) probability measure on the process remains the same, \( P_s, \pi, \rho = P_s, \pi^*, \rho \) (for a proof, see the author's master's thesis [3]).

Now because of the weakening of condition \( h \), the function \( v \) found in theorem 3.6 represents the value of the game within the class \( R \) only. However, by applying the above-mentioned theorem, we see that \( v \) is also the value of the game on \( \Pi \)

\[
(\forall \in \Pi(B) \Pi(A) \leq \inf \sup \nu(\pi, \rho) = \inf \sup v(\pi, \rho) \leq \sup \inf v(\pi, \rho) \leq v^L). \]

In the same way theorem 3.8 is proved first for Markov strategies \( (\pi, \rho) \); then the transition to general strategies can be made.

This way of proving may also be applied to theorem 4.1, so that evidently
in condition $h$ II can be replaced by $R$.

**Remark 5.3.** If we take $\mu = e (\mu(s) = 1)$ in the previous sections, then with $\beta' = 1$ conditions $b$ and $f$ are satisfied. If the reward $r$ is bounded w.r.t. $\mu$ and $\beta \beta' < 1$, then $g$, $g$ and $h$ are fulfilled. Now it is easy to see that theorem 5.1 is a generalization of Parthasarathy's Theorem 3.1 [8], and theorem 4.1 a generalization of his Theorem 3.2.

**Remark 5.4.** In the author's master's thesis [3] assumption $f$ of theorem 3.1 is closer investigated, and in particular the stronger assumption $f^*$: there exists a $\beta' \in (0,1)$ such that for all $s$, $a$ and $b$

$$\int \frac{p(ds'|s,a,b)}{\mu(s')} \leq \beta' \mu(s).$$

The Markov processes $\{X_t \mid t = 0,1,\ldots\}$ ($X_t$ is the state at time $t$) generated by choosing a starting state $s \in S$ and a pair of Markov strategies $(\pi, \rho) \in R$ are said to be contracting if the transition law $p$ satisfies $f^*$, for some measurable $\mu : S \to (0,\infty)$. Now if for all $s$, $a$ and $b$ $p(\cdot |s,a,b)$ is absolutely continuous with respect to a fixed $\sigma$-finite measure $q$ on $S$, two characterizations can be given of this contraction property (actually of a somewhat weaker property, "$q$-contraction"): one in the form of an enforced drift through a partitioning of the state space, and one in terms of an exponentially bounded lifetime (these results are a generalization of corresponding results for countable $S$, presented by Van Hee and Wessel [4]).

**References**


