Characterization of strong (Nash) equilibrium points in Markov games

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Mathematics

PROBABILITY THEORY, STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 77-09

Characterization of strong (Nash) equilibrium points in Markov games

by

H.A.M. Couwenbergh

Eindhoven, April 1977

The Netherlands

Characterization of strong (Nash) equilibrium points in Markov games

by

H.A.M. Couwenbergh

Summary

The main result in this paper is the characterization of certain strong kinds of equilibrium points in Markov games with a countable set of players and uncountable decision sets. Two person Markov games are studied beforehand, since this paper gives an extension of the existing theory for two person zero sum Markov games; finally we consider the special cases of N-person Markov games and Markov decision processes.

1. Introduction

This paper describes the results obtained in an attempt to extend the theory concerning optimal strategies in two person zero sum Markov games, as developed by Groenewegen in [1]; the extension being directed towards general (more persons) Markov games with individual rewards, where (Nash) equilibrium points are the equivalent of optimal strategies.
In advance an important observation: in [1] the possibility of defining a fixed optimal value v underlies the definitions of the basic concepts: in our case no such value exists (generally); it will be shown, nevertheless, that it is sufficient assuming an arbitrary, if necessary time-dependent value v_n for player n, with respect to which we can work instead.
As is done in [1], we try to characterize an equilibrium point (abbreviation: eqpt) satisfying some extra conditions by the combination of two concepts:

i) a sort of policy equilibrium property holding for all points of time, called saddlingness; and
ii) an "asymptotic definiteness" property, which prevents that a player ultimately receives more than he could expect, whatever strategy he chooses himself.

It turns out that such a characterization can be found for two kinds of equilibrium points: a strong kind coinciding with the set of "subgame perfect" strategies (see [1]) for the two person zero sum case, and a weaker kind called semi-persistent, which consists of eqpt^S in that special case standing midway between "persistently optimal" ([1]) and subgame perfect strategies (subsections 2.1, 2.2, 2.3 and 2.4). In [1] it is demonstrated that under certain "equalizing" conditions every optimal strategy can be improved to a subgame perfect one. However, the method used there is not (at least not easily) adaptable in order to transform the "ordinary" eqpt^S of two person non-zero sum Markov games into eqpt^S belonging to one of the above-mentioned kinds (subsection 2.5).

As for Markov games with countable set of players, we can now describe the analogues of the two mentioned kinds in a simple way (a more subtle framework is required, though): this is due to the fact that already in the two person case the definitions, theorems and proofs fall apart into two independent parts (viz. separately for the rewards of player A and the rewards of player B) (subsections 3.1 and 3.2).

N-person Markov games can be treated as a special case; finally we view the results for N = 1, actually Markov decision processes (subsection 3.3).

2. Two person Markov games

2.1. The model

The players are called A and B and the game is described by

S : the countable or finite state space;

$K = \underset{i \in S}{\times} K_i$ (Cartesian product): the action space for player A, K_i being the set of actions available, with K_i countable or finite and nonempty, for all $i \in S$;

$L = \underset{i \in S}{\times} L_i$: action space for player B (analogous to K);

p : a function $\{(i,j,k,\ell) \mid i,j \in S,\ k \in K_i,\ \ell \in L_i\} \to [0,1]$ such that $p(i,j,k,\ell)$ represents the probability of reaching state j after one unit of time, given the present state i and the actions k and ℓ taken there by A and B respectively; we do not require $\sum_{j \in S} p(i,j,k,\ell) = 1$, only $\sum_{j \in S} p(i,j,k,\ell) \leq 1$;

r_1: a function $\{(i,k,\ell) \mid i \in S,\ k \in K_i,\ \ell \in L_i\} \to \mathbb{R}$, $r_1(i,k,\ell)$ being the reward player A receives in state i when the actions k and ℓ are chosen;

r_2: a similar function as reward for B.

Let for $i \in S$ $F(i)$ and $G(i)$ be the set of randomized actions (*policies*) over K_i and L_i respectively; $F := \times_{i \in S} F(i)$, $G := \times_{i \in S} G(i)$; for $f \in F$ we denote by $f(i,k)$ the probability of A taking action $k \in K_i$ when the current state is i, similarly for player B. Define

$$\forall_{f\in F}\ \forall_{g\in G}\ \forall_{i,j\in S}\ [r_1(i,f,g) := \sum_{k,\ell} r_1(i,k,\ell)f(i,k)g(i,\ell),$$

$$r_2(i,f,g) := \sum_{k,\ell} r_2(i,k,\ell)f(i,k)g(i,\ell),$$

$$(P(f,g))_{ij} := \sum_{k,\ell} p(i,j,k,\ell)f(i,k)g(i,\ell)] ;$$

we assume absolute convergence of all these sums; omitting the indices i and j we denote by $r_1(f,g)$ ($r_2(f,g)$) the vector with components $r_1(i,f,g)$ ($r_2(i,f,g)$), and by $P(f,g)$ the matrix with components $(P(f,g))_{ij}$.
The time a transition requires equals unity, the starting time is 0.

A Markov strategy for player A is a sequence $\pi = (f_0,f_1,\ldots)$ with $\forall_{t\geq 0}[f_t \in F]$. The set of all strategies of this kind is called R(A); similarly for strategy $\rho = (g_0,g_1,\ldots) \in R(B)$, $\forall_t\ [g_t \in G]$; $R := R(A) \times R(B)$. Conversely, if $(\pi,\rho) \in R$, then we indicate the components of π by f_t, those of ρ by g_t, $t = 0,1,2,\ldots$.

A strategy $(\pi,\rho) \in R$ and a starting state $i \in S$ determine a stochastic process X_t ($t = 0,1,\ldots$) on S (X_t is the state of the system at time t): that is, in period $(t,t+1)$ the system moves according to the transition matrix $P(f_t,g_t)$. The probability measure for this process will be denoted by $P_i^{(\pi,\rho)}$; $E_i^{(\pi,\rho)}$ represents the corresponding expectation operator.
In the remainder of section 2 we shall be working under assumption A (*charge structure*): for all $(\pi,\rho) \in R$ and $i \in S$

$$w_1(i,\pi,\rho) := E_i^{(\pi,\rho)} \sum_{t=0}^{\infty} |r_1(X_t,f_t,g_t)| < \infty ,$$

$$w_2(i,\pi,\rho) := E_i^{(\pi,\rho)} \sum_{t=0}^{\infty} |r_2(X_t,f_t,g_t)| < \infty .$$

Now we are able to define

$$v_1(i,\pi,\rho) := E_i^{(\pi,\rho)} \sum_{t=0}^{\infty} r_1(X_t,f_t,g_t) ,$$

$$v_2(i,\pi,\rho) := E_i^{(\pi,\rho)} \sum_{t=0}^{\infty} r_2(X_t,f_t,g_t) \quad \text{for all } (\pi,\rho) \in R \text{ and } i \in S .$$

Let $(\pi,\rho) = (f_0,f_1,\ldots,g_0,g_1,\ldots) \in R$ and $t \geq 0$ be arbitrary, then

$$\pi(t) := (f_t,f_{t+1},\ldots) ;$$

$$\rho(t) := (g_t,g_{t+1},\ldots) ;$$

$$S^{(\pi,\rho)}(t) := \{j \in S \mid \exists_{i\in S} [P_i^{(\pi,\rho)}(X_t = j) > 0]\}$$

(the set of those $j \in S$ that can be reached at time t; we have $S^{(\pi,\rho)}(0) = S$);

$$S(B,\rho,t) := \{j \in S \mid \exists_{i\in S} \exists_{\pi'\in R(A)} [P_i^{(\pi',\rho)}(X_t = j) > 0]\} ;$$

$$S(A,\pi,t) := \{j \in S \mid \exists_{i\in S} \exists_{\rho'\in R(B)} [P_i^{(\pi,\rho')}(X_t = j) > 0]\} ;$$

$$S(t) := \{j \in S \mid \exists_{i\in S} \exists_{(\pi',\rho')\in R} [P_i^{(\pi',\rho')}(X_t = j) > 0]\} .$$

Obviously

$$S^{(\pi,\rho)}(t) \subset S(B,\rho,t) \cap S(A,\pi,t)$$

and

$$S(B,\rho,t) \cup S(A,\pi,t) \subset S(t) .$$

2.2. Equilibrium points

A strategy $(\pi,\rho) \in R$ is called an equilibrium point in $j \in S$ iff

$$\forall_{(\pi',\rho')\in R} [v_1(j,\pi,\rho) \geq v_1(j,\pi',\rho) \text{ and } v_2(j,\pi,\rho) \geq v_2(j,\pi,\rho')] ;$$

(π,ρ) is called a (Nash) equilibrium point iff this statement holds for all $j \in S$.

We derive two properties of eqpt^S, the first one condensed in

(2.1.1) Lemma. $(\pi,\rho)\text{eqpt} \Rightarrow \forall_t \forall_{j\in S^{(\pi,\rho)}(t)} [(\pi(t),\rho(t))\text{eqpt in } j]$.

Proof. Let $t \geq 0$ and $j \in S^{(\pi,\rho)}(t)$, so for some $i \in S$ we have $P_i^{(\pi,\rho)}(X_t = j) > 0$. In this proof we use non-Markov strategies; a model for two person nonzero sum games with general strategies (these may depend on the history) can be constructed as an extension of [2] paragraph 1: we call $\Pi(A)$ the set of all strategies for player A, $\Pi(B)$ for B, $v_1(i,\pi,\rho)$ the expected total reward for A (expectation with respect to the prob. measure $P_i^{(\pi,\rho)}$ for the generalized strategies) when the starting state is $i \in S$, and $(\pi,\rho) \in \Pi(A) \times \Pi(B)$, etc.

We have

$$(*) \qquad \forall_{\pi^* \in R(A)} [v_1(i,\pi,\rho) \geq v_1(i,\pi^*,\rho)] \Rightarrow \forall_{\pi^* \in \Pi(A)} [v_1(i,\pi,\rho) \geq v_1(i,\pi^*,\rho)]:$$

this follows from [2] lemma 1.

Choose $\pi' \in R(A)$; we construct the (non-Markov) strategy π^0 as follows: let π^0 be equal to π, except that if $X_t = j$, then π^0 is from time t on the same as π'.

Since (π,ρ) is an eqpt, the left-hand part of (*) holds, so we have $v_1(i,\pi,\rho) \geq v_1(i,\pi^0,\rho)$. Consequently,

$$\sum_{s=0}^{t-1} E_i^{(\pi,\rho)} r_1(X_s,f_s,g_s) + \sum_{\ell \in S} P_i^{(\pi,\rho)}(X_t = \ell) v_1(\ell,\pi(t),\rho(t)) =$$

$$= v_1(i,\pi,\rho) \geq v_1(i,\pi^0,\rho) = \sum_{s=0}^{t-1} E_i^{(\pi,\rho)} r_1(X_s,f_s,g_s) +$$

$$+ \sum_{\ell \in S, \ell \neq j} P_i^{(\pi,\rho)}(X_t = \ell) v_1(\ell,\pi(t),\rho(t)) + P_i^{(\pi,\rho)}(X_t = j) v_1(j,\pi',\rho(t)),$$

from which we obtain

$$v_1(j,\pi(t),\rho(t)) \geq v_1(j,\pi',\rho(t)) \ .$$

Analogously

$$\forall_{\rho' \in R(B)} [v_2(j,\pi(t),\rho(t)) \geq v_2(j,\pi(t),\rho')] \ ;$$

so $(\pi(t),\rho(t))$ is an eqpt in j. □

By means of lemma (2.1.1) we deduce

(2.1.2) Theorem. If $(\pi,\rho) \in R$ is an equilibrium point, then

$$\forall_t \forall_{j \in S^{(\pi,\rho)}(t)} \forall_{f \in F} [v_1(j,\pi(t),\rho(t)) \geq r_1(j,f,g_t) + (P(f,g_t)v_1(\pi(t+1),\rho(t+1)))(j)],$$

$$\forall_t \forall_{j \in S^{(\pi,\rho)}(t)} \forall_{g \in G} [v_2(j,\pi(t),\rho(t)) \geq r_2(j,f_t,g) + (P(f_t,g)v_2(\pi(t+1),\rho(t+1)))(j)].$$

Proof.

i) Let $t \geq 0$, $j \in S^{(\pi,\rho)}(t)$ and $f \in F$, then according to (2.1.1)

$$v_1(j,\pi(t),\rho(t)) \geq v_1(j,f\pi(t+1),\rho(t)) =$$

$$= r_1(j,f,g_t) + (P(f,g_t)v_1(\pi(t+1),\rho(t+1)))(j) \ ;$$

ii) analogously for v_2. □

Interpretation of (2.1.2): at any time t the tails of π and ρ, $\pi(t)$ and $\rho(t)$, prescribe in the (under (π,ρ)) attainable states policies f_t and g_t, so that a player cannot gain anything by taking an other policy at time t.
The "self-saddle-conserving" property expressed in (2.1.2) is not a sufficient condition for an equilibrium point: even in zero sum games this is not the case (see [1], here an eqpt(π,ρ) is an optimal strategy; and the property (2.1.2) is weaker than "saddle conserving" as defined in [1]).
As is done in [1], we therefore consider, in the next subsection, strategies satisfying certain stronger demands (than "common" eqpt^s): the purpose being to give an exact characterization (necessary and sufficient conditions) of such strategies by a property similar to (2.1.2); also an "asymptotic definiteness" property is needed.

2.3. Characterization of (v_1,v_2)-semi-persistent and (v_1,v_2)-subgame perfect equilibrium points

In this subsection we need two functions v_1 and v_2: $S \times (\mathbb{N} \cup \{0\}) \to \mathbb{R}$; unless these functions are specified below, we assume that they are given; v_1 and v_2 must satisfy
Assumption B: $\forall_t \forall_{f\in F} \forall_{g\in G}$ [$P(f,g)|v_1(t)| < \infty$ and $P(f,g)|v_2(t)| < \infty$ in all components] (by $|v_1(t)|$ is meant the vector $\{|v_1(j,t)|\}_{j\in S}$).

B is fulfilled for instance if $\forall_t \exists_M \forall_{j\in S}$ [$|v_1(j,t)| \le M$ and $|v_2(j,t)| \le M$]; B is also satisfied when we choose some $(\pi,\rho) \in R$ and define

$$\forall_t \forall_j [v_1(j,t) := v_1(j,\pi(t),\rho(t)), v_2(j,t) := v_2(j,\pi(t),\rho(t))] :$$

viz.

$$|v_1(i,t)| = |v_1(i,\pi(t),\rho(t))| \le w_1(i,\pi(t),\rho(t)),$$

so

$$(P(f,g)|v_1(t)|)(j) \le (P(f,g)w_1(\pi(t),\rho(t)))(j) + |r_1(j,f,g)| =$$

$$= w_1(j,f\pi(t),g\rho(t)) < \infty$$

owing to assumption A.
We introduce three concepts and then prove that the second and third together characterize the first.

$(\pi,\rho) \in R$ is called $\underline{(v_1,v_2)\text{-semi-persistent}}$ iff

$$(2.3.1)\quad \begin{cases} \forall_t \forall_{j\in S(B,\rho,t)} \forall_{\pi'\in R(A)} \left[v_1(j,\pi(t),\rho(t)) = v_1(j,t) \geq v_1(j,\pi',\rho(t))\right] \\ \forall_t \forall_{j\in S(A,\pi,t)} \forall_{\rho'\in R(B)} \left[v_2(j,\pi(t),\rho(t)) = v_2(j,t) \geq v_2(j,\pi(t),\rho')\right]; \end{cases}$$

$(\pi,\rho) \in R$ is called $\underline{(v_1,v_2)\text{-saddling}}$ iff

$$(2.3.2)\quad \begin{cases} \forall_t \forall_{j\in S(B,\rho,t)} \forall_{f\in F} \left[r_1(j,f_t,g_t) + (P(f_t,g_t)v_1(t+1))(j) = v_1(j,t) \geq \right. \\ \qquad\qquad \left. \geq r_1(j,f,g_t) + (P(f,g_t)v_1(t+1))(j)\right] \\ \forall_t \forall_{j\in S(A,\pi,t)} \forall_{g\in G} \left[r_2(j,f_t,g_t) + (P(f_t,g_t)v_2(t+1))(j) = v_2(j,t) \geq \right. \\ \qquad\qquad \left. \geq r_2(j,f_t,g) + (P(f_t,g)v_2(t+1))(j)\right]; \end{cases}$$

$(\pi,\rho) \in R$ is called $\underline{(v_1,v_2)\text{-asymptotically definite}}$ iff

$$(2.3.3)\quad \begin{cases} \forall_t \forall_{j\in S(B,\rho,t)} \forall_{\pi'\in R(A)} \left[\lim_{k\to\infty} E_j^{(\pi(t),\rho(t))} v_1(X_k,t+k) = 0 = \right. \\ \qquad\qquad \left. \lim_{k\to\infty} E_j^{(\pi',\rho(t))} v_1^-(X_k,t+k)\right] \\ \forall_t \forall_{j\in S(A,\pi,t)} \forall_{\rho'\in R(B)} \left[\lim_{k\to\infty} E_j^{(\pi(t),\rho(t))} v_2(X_k,t+k) = 0 = \right. \\ \qquad\qquad \left. \lim_{k\to\infty} E_j^{(\pi(t),\rho')} v_2^-(X_k,t+k)\right], \end{cases}$$

where $a^- := \max\{0,-a\}$.

It is clear that [(π,ρ) is (v_1,v_2)-semi-persistent] $\Rightarrow$ [(π,ρ) is eqpt]; for an interpretation of saddlingness and asymptotic definiteness see the Introduction.

(2.3.4) $\underline{\text{Theorem.}}$ If $(\pi,\rho) \in R$, then (π,ρ) is (v_1,v_2)-semi-persistent iff (π,ρ) is (v_1,v_2)-saddling and (v_1,v_2)-asymptotically definite.

$\underline{\text{Proof.}}$ $\Rightarrow$: i) Choose $t \geq 0$, $j \in S(B,\rho,t)$ and $f \in F$. Then $v_1(j,t) = v_1(j,\pi(t),\rho(t))$ $= r_1(j,f_t,g_t) + (P(f_t,g_t)v_1(\pi(t+1),\rho(t+1)))(j) = r_1(j,f_t,g_t) + (P(f_t,g_t)v_1(t+1))(j)$ since for all i satisfying $(P(f_t,g_t))_{ji} > 0$ we have $i \in S(B,\rho,t+1)$, so $v_1(i,\pi(t+1),\rho(t+1)) = v_1(i,t+1)$. A similar reasoning gives $v_1(j,t) \geq$ $\geq v_1(j,f\pi(t+1),\rho(t)) = r_1(j,f,g_t) + (P(f,g_t)v_1(t+1))(j)$. Likewise we treat v_2. Conclusion: (π,ρ) is (v_1,v_2)-saddling.

ii) Now the asymptotic definiteness: let $t \geq 0$, $j \in S(B,\rho,t)$ and $\pi' \in R(A)$. We have $|E_j^{(\pi(t),\rho(t))} v_1(X_k,t+k)| \leq E_j^{(\pi(t),\rho(t))} |v_1(X_k,t+k)| =$
$= E_j^{(\pi(t),\rho(t))} |v_1(X_k,\pi(t+k),\rho(t+k))| \leq E_j^{(\pi(t),\rho(t))} \sum_{m=k}^{\infty} |r_1(X_m,f_{t+m},g_{t+m})| \to 0$
$(k \to \infty)$ on account of A.
Using $a \geq b \Rightarrow a^- \leq |b|$, we can write

$$0 \leq E_j^{(\pi',\rho(t))} v_1^-(X_k,t+k) \leq E_j^{(\pi',\rho(t))} |v_1(X_k,\pi'(k),\rho(t+k))| \to 0 \quad (k \to \infty)$$

on grounds similar to those above.
In the same way we handle v_2. So (π,ρ) is (v_1,v_2)-asymptotically definite.
$\Leftarrow$: Assume $t \geq 0$, $j \in S(B,\rho,t)$ and $\pi' = (f_0',f_1',\ldots) \in R(A)$ to be given; now

$$v_1(j,t) = r_1(j,f_t,g_t) + (P(f_t,g_t)v_1(t+1))(j) = r_1(j,f_t,g_t) +$$

$$(P(f_t,g_t)r_1(f_{t+1},g_{t+1}))(j) + (P(f_t,g_t)P(f_{t+1},g_{t+1})v_1(t+2))(j) = \ldots =$$

$$E_j^{(\pi(t),\rho(t))} \sum_{m=0}^{\infty} r_1(X_m,f_{t+m},g_{t+m}) + \lim_{k\to\infty} E_j^{(\pi(t),\rho(t))} v_1(X_k,t+k) =$$

$$v_1(j,\pi(t),\rho(t)) \ ;$$

and

$$v_1(j,t) \geq r_1(j,f_0',g_t) + (P(f_0',g_t)v_1(t+1))(j) \geq \ldots \geq$$

$$v_1(j,\pi',\rho(t)) - \lim_{k\to\infty} E_j^{(\pi',\rho(t))} v_1^-(X_k,t+k) = v_1(j,\pi',\rho(t)) \ .$$

Analogous proof for v_2. Consequently, (π,ρ) is (v_1,v_2)-semi-persistent. □

This procedure can be repeated for the three stronger concepts introduced in the following way:
If $(\pi,\rho) \in R$ satisfies (2.3.1), (2.3.2) or (2.3.3) with $S(B,\rho,t)$ and $S(A,\pi,t)$ replaced by $S(t)$, (π,ρ) is called respectively (v_1,v_2)-subgame perfect (originally introduced by Selten in [5]), (v_1,v_2)-overall saddling or (v_1,v_2)-overall asymptotically definite.
Naturally (v_1,v_2)-subgame perfect is stronger than (v_1,v_2)-semi-persistent.
Now the characterization of (v_1,v_2)-subgame perfect:

(2.3.5) Theorem. If $(\pi,\rho) \in R$, then (π,ρ) is (v_1,v_2)-subgame perfect iff (π,ρ) is (v_1,v_2)-overall saddling and (v_1,v_2)-overall asymptotically definite.

Proof. Entirely similar to the proof of (2.3.4). □

Let $(\pi,\rho) \in R$. We can take

$$\forall_{t\geq 0}\ \forall_{j\in S}\ [v_1(j,t) := v_1(j,\pi(t),\rho(t)),\ v_2(j,t) := v_2(j,\pi(t),\rho(t))];$$

if in this particular case (π,ρ) satisfies (2.3.1), (2.3.2) or (2.3.3), we call (π,ρ) semi-persistent, saddling or asymptotically definite respectively. According to (2.3.4) we have

(π,ρ) semi-persistent $\Leftrightarrow$ (π,ρ) saddling and asymptotically definite

(notice that the left part of the statements in (2.3.1), (2.3.2) and (2.3.3) is now trivial).
The same v_1 and v_2 can be used in theorem (2.3.5); here "(π,ρ) is subgame perfect" is equivalent to $\forall_t\ \forall_{j\in S(t)}\ [(\pi(t),\rho(t))$ eqpt in j].

2.4. Specialization: zero sum games

In the special case of zero sum Markov games we have $r_1 = -r_2 =: r$ and define $v_1 := -v_2 := v := \sup_{\pi\in R(A)} \inf_{\rho\in R(B)} v(\pi,\rho)$ (supinf componentwise), so that v_1 and v_2 do not depend on the time-variable. In [2] it is established that, under an assumption slightly stronger than A, we have $|v| < \infty$ and $\forall_{f,g}\ [P(f,g)|v| < \infty]$, so B is satisfied.

As for the stronger kind of eqpt^S (now equivalent to optimal strategies), it is readily checked that (v_1,v_2)-subgame perfect, (v_1,v_2)-overall saddling and (v_1,v_2)-overall asymptotically definite are the same as the following concepts originating from [1]: respectively subgame perfect, overall saddling and overall asymptotically definite. Hence Theorem 6.1 from [1] is a consequence of (2.3.5).

We show now that the property (v,-v)-semi-persistent, shortly v-semi-persistent, lies between subgame perfect and persistently optimal (- the last being defined in [1] as follows: $(\pi,\rho) \in R$ is called persistently optimal iff for all $t \geq 0$ $\forall_{j\in S(B,\rho,t)}\ [(\pi,\rho(t))$ optimal in j] and $\forall_{j\in S(A,\pi,t)}\ [(\pi(t),\rho)$ optimal in j]):

(2.4.1) Theorem. If $(\pi,\rho) \in R$ then

i) (π,ρ) subgame perfect $\Rightarrow$ (π,ρ) v-semi-persistent;

ii) (π,ρ) v-semi-persistent $\Rightarrow$ (π,ρ) persistently optimal (in this special zero sum case our nomenclature seems a bit odd).

Proof.

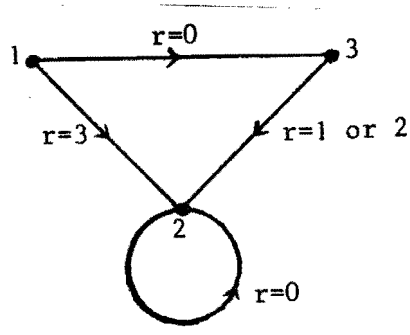
i) This implication is obvious (subgame perfect = (v,-v)-subgame perfect ⇒ (v,-v)-semi-persistent = v-semi-persistent).

ii) Choose $t \geq 0$ and $j \in S(B,\rho,t)$. We have to prove that $(\pi,\rho(t))$ is optimal in j. Let $(\pi',\rho') \in R$. From (π,ρ) v-semi-persistent it follows that (π,ρ) is an eqpt, which is the same as (π,ρ) optimal, so $v(j) = v(j,\pi,\rho) \leq v(j,\pi,\rho')$ At the same time, according to (2.3.1), $v(j) \geq v(j,\pi',\rho(t))$; by taking for one moment $\rho' = \rho(t)$ and $\pi' = \pi$, we see that $v(j) = v(j,\pi,\rho(t))$. Concluding: for all $(\pi',\rho') \in R$ we have $v(j,\pi',\rho(t)) \leq v(j) = v(j,\pi,\rho(t)) \leq$ $\leq v(j,\pi,\rho')$, so $(\pi,\rho(t))$ is optimal in j. We can prove in a similar way that $\forall_t \forall_{j \in S(A,\pi,t)} [(\pi(t),\rho)$ eqpt in j]. Consequently (π,ρ) is persistently optimal. □

Persistent optimality is not the same as v-semi-persistency:

(2.4.2) Counterexample with (π,ρ) persistently optimal and not v-semi-persistent. We have $S = \{1,2,3\}$, $K_1 = \{a,b\}$, $K_2 = \{0\}$, $K_3 = \{c,d\}$, $L_1 = L_2 = L_3 = \{0\}$ (so player B has nothing to decide upon, actually). Omitting the indices for player B (e.g. p(1,2,a,0) becomes p(1,2,a)), we define p(1,2,a) = p(1,3,b) = = p(2,2,0) = p(3,2,c) = p(3,2,d) = 1, the other transition probabilities as zero; further r(1,a) = 3, r(1,b) = 0 = r(2,0), r(3,c) = 1 and r(3,d) = 2. Now v(1) = 3, v(2) = 0, v(3) = 2. Take

$$\pi := \begin{pmatrix} a & a & a & \dots \\ 0 & 0 & 0 & \dots \\ d & c & c & \dots \end{pmatrix}, \quad \rho := \begin{pmatrix} 0 & 0 & \dots \\ 0 & 0 & \dots \\ 0 & 0 & \dots \end{pmatrix}$$



(only nonrandomized policies, first row for state 1, second for state 2 and third for state 3; the first column represents the policy for t = 0 and so on).
It is easy to verify that (π,ρ) is persistently optimal; however (π,ρ) is not v-semi-persistent: $3 \in S(B,\rho,1)$ but $v(3) = 2 \neq 1 = v(3,\pi(1),\rho(1))$. □

The point is, that if $j \in S(B,\rho,t) \setminus S(A,\pi,t)$, persistent optimality cannot prohibit a "bad" tail $\pi(t)$ in j, whereas v-semi-persistency prevents incidents of this kind.

Remark. No suitable extension of persistent optimality to nonzero sum Markov games was found: this concept is apparently too weak for a characterization similar to (2.3.4).

2.5. Improving ordinary equilibrium points

Several theorems in literature guarantee, under fairly general conditions, the existence of eqpt^S (see e.g. [4]: two person Markov games with discount factor < 1 possess even a stationary eqpt if S, K and L are finite). These are "ordinary" eqpt^S, that is, not necessarily semi-persistent. In [1] the question of the existence of subgame perfect strategies is reduced to the question of the existence of ordinary eqpt^S, by indicating a method with which any optimal strategy may be improved to overall saddling (and this is the same as subgame perfect under certain "equalizing" conditions). This method however cannot be used in general two person Markov games, as is shown below.

The method developed in [1] runs as follows: suppose (π,ρ) is eqpt, then define (π^*,ρ^*) by

$$f_0^* := f_0,\ g_0^* := g_0;$$

$$t > 0:\ \begin{cases} j \in S^{(\pi,\rho)}(t):\ f_t^*(j,\cdot) := f_t(j,\cdot),\ g_t^*(j,\cdot) := g_t(j,\cdot)\ ; \\ j \notin S^{(\pi,\rho)}(t):\ f_t^*(j,\cdot) := f_{t-1}^*(j,\cdot),\ g_t^*(j,\cdot) := g_{t-1}^*(j,\cdot). \end{cases} \tag{2.5.1}$$

In case $r_1 = -r_2$, we have: (π^*,ρ^*) is saddling if (π,ρ) is "saddle conserving" ([2] lemma 6). For nonzero sum games we have the following

(2.5.2) Counterexample with (π,ρ) subgame perfect, yet (π^*,ρ^*) not even an eqpt.

S = {1,2,3,4,5}, discounting with factor ½ (this may be fitted in the original model (2.1) by adding an absorbing extra state with reward 0 for both players, where the state variable X_t arrives with probability ½ for all $t \geq 1$); $K_1 = \{x,y\}$, $K_2 = \{a,b\}$, $L_1 = \{0\}$, $L_2 = \{a,b\}$, in the absorbing states 3, 4 and 5 no choice of decision is possible; all transition probabilities are equal to zero or one, see the figure and the following matrices:

transition matrix in state 1:

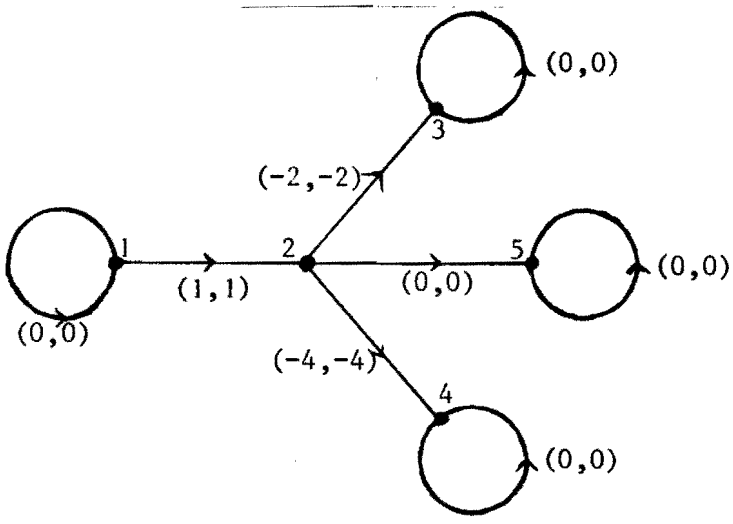
K_1 \ L_1	0
x	1
y	2

; in state 2:

K_2 \ L_2	a	b
a	5	4
b	4	3

(this means that for instance p(2,4,b,a) = 1).

The pair (r_1,r_2) by a transition in the figure indicates the rewards for A and B respectively, corresponding to this transition.



Take

$$\binom{\pi}{\rho} := \begin{pmatrix} x & x & x & x & \ldots \\ a & b & b & b & \ldots \\ a & b & b & b & \ldots \end{pmatrix}$$

(nonrandomized policies, the first row gives the successive actions for A in state 1, the second row the actions for A in state 2, and the third the actions for B in state 2, the actions for B in state 1 are omitted).

It is easily checked that (π,ρ) satisfies (2.3.1) with $S(t)$ instead of $S(B,\rho,t)$ and $S(A,\pi,t)$, and with $v_1(j,t) = v_1(j,\pi(t),\rho(t))$, $v_2(j,t) = v_2(j,\pi(t),\rho(t))$ (e.g. $v_1(1,\pi,\rho) = 0 \geq v_1(1,\pi',\rho)$ for all $\pi' \in R(A)$, since the system remains in state 1 if it is started there; and π' instead of π can never anymore generate the profitable combination $\binom{a}{a}$ for state 2).
We construct the "improved" strategy in accordance with (2.5.1):

$$\binom{\pi^*}{\rho^*} = \begin{pmatrix} x & x & x & \ldots \\ a & a & a & \ldots \\ a & a & a & \ldots \end{pmatrix}$$

(observe $t > 0 \Rightarrow 2 \notin S^{(\pi,\rho)}(t)$). Now (π^*,ρ^*) is not an eqpt:

$$v_1(1,\pi^*,\rho^*) = 0 \not\geq 1 + \tfrac{1}{2}.0 = v_1(1,\binom{y}{a}\pi^*(1),\rho^*) \ . \qquad \square$$

Other obvious "improving" methods fail too, such as ((π,ρ) a given eqpt)

$$f_0^* := f_0,\ g_0^* := g_0\ ;$$

(2.5.3)
$$t>0:\begin{cases} j\in S(B,\rho,t)\cup S(A,\pi,t):\ f_t^*(j,\cdot) := f_t(j,\cdot),\ g_t^*(j,\cdot) := g_t(j,\cdot);\\ j\notin S(B,\rho,t)\cup S(A,\pi,t):\ f_t^*(j,\cdot) := f_{t-1}^*(j,\cdot),\ g_t^*(j,\cdot) := g_{t-1}^*(j,\cdot).\end{cases}$$

(2.5.4) Counterexample where the (π^*,ρ^*) obtained from an eqpt (π,ρ) according to (2.5.3) is not semi-persistent:
we can take the same model as described in (2.5.2) and define

$$\binom{\pi}{\rho} := \begin{pmatrix} x & x & x & x & x & \dots\\ a & b & a & b & b & \dots\\ a & b & b & b & b & \dots\end{pmatrix}.$$

This is an equilibrium point; yet $(\pi^*,\rho^*) = (\pi,\rho)$ is not semi-persistent:
$2\in S(B,\rho^*,2)$, but

$$v_1(2,\pi^*(2),\rho^*(2)) = -4 \not\geq -2 = v_1(2,\binom{x}{b}\pi^*(3),\rho^*(2))\ .$$ □

Or

$$f_0^* := f_0,\ g_0^* := g_0\ ;$$

(2.5.4)
$$t>0:\begin{cases} j\in S(B,\rho,t):\ f_t^*(j,\cdot) := f_t(j,\cdot),\ j\notin S(B,\rho,t):\ f_t^*(j,\cdot) := f_{t-1}^*(j,\cdot);\\ j\in S(A,\pi,t):\ g_t^*(j,\cdot) := g_t(j,\cdot),\ j\notin S(A,\pi,t):\ g_t^*(j,\cdot) := g_{t-1}^*(j,\cdot):\end{cases}$$

(2.5.5) Counterexample: as (2.5.2), with the same (π,ρ) we get

$$\binom{\pi^*}{\rho^*} = \begin{pmatrix} x & x & x & \dots\\ a & b & b & \dots\\ a & a & a & \dots\end{pmatrix}: \text{ not semi-persistent.}$$ □

3. Markov games with countable set of players

1. Model

The players are numbered 1,2,...; the state space S is again finite or countably infinite. For player n and state $i\in S$ let $K_i^{(n)}$ be the (not necessarily countable) action space, $\Gamma_i^{(n)}$ a σ-algebra of subsets of $K_i^{(n)}$, and $M_i^{(n)}$ a set of probability measures on $(K_i^{(n)},\Gamma_i^{(n)})$; $M^{(n)} := \underset{i\in S}{\times} M_i^{(n)}$, that is, if $\mu^{(n)}\in M^{(n)}$, then $\mu^{(n)}(i,\cdot)$ is a probability distribution over $(K_i^{(n)},\Gamma_i^{(n)})$ for every $i\in S$.

$K_i := \overset{\infty}{\underset{n=1}{\times}} K_i^{(n)}$, Γ_i is defined as the product σ-algebra on K_i generated by $\Gamma_i^{(1)}, \Gamma_i^{(2)}, \ldots$, this for all $i \in S$.
When all players have chosen a policy, namely $\forall_{n\in\mathbb{N}} [\mu^{(n)} \in M^{(n)}]$, we define $\underline{\mu} := (\mu^{(1)}, \mu^{(2)}, \ldots)$: this is the simultaneous policy, an element of $M := \overset{\infty}{\underset{n=1}{\times}} M^{(n)}$ (simultaneous decisions are underlined); now $\mu(i,\cdot)$ denotes the infinite product measure $\mu^{(1)}(i,\cdot) \otimes \mu^{(2)}(i,\cdot) \otimes \ldots$ on (K_i, Γ_i) (generated therefore by $\mu^{(1)}(i,\cdot)$, $\mu^{(2)}(i,\cdot)$ and so on).

We assume that a function $p := \{(i,j,\underline{k}) \mid i,j \in S, \underline{k} \in K_i\} \to [0,1]$ is given with the property that for all $i,j \in S$ $p(i,j,\cdot)$ is a measurable function with respect to Γ_i, and so that $p(i,j,\underline{k})$ is the probability of the transition from i to j if the (simultaneous) action $\underline{k} \in K_i$ is taken ($\sum_{j\in S} p(i,j,\underline{k}) \leq 1$); also the functions r_n: $\{(i,\underline{k}) \mid i \in S, \underline{k} \in K_i\} \to \mathbb{R}$ ($n \in \mathbb{N}$) are given, the rewards for the players, where it is assumed that for all $i \in S$ $r_n(i,\cdot)$ is Γ_i-measurable. Define

$$\forall_{\underline{\mu}\in M} \forall_{i,j\in S} \left[(P(\underline{\mu}))_{ij} := \int_{K_i} p(i,j,\underline{k}) d\mu(i,\underline{k})\right],$$

$$\forall_{n\in\mathbb{N}} \forall_{\underline{\mu}\in M} \forall_{i\in S} \left[r_n(i,\underline{\mu}) := \int_{K_i} r_n(i,\underline{k}) d\mu(i,\underline{k})\right];$$

we assume the absolute convergence of all these integrals. $P(\underline{\mu})$ and $r_n(\underline{\mu})$ are the corresponding matrix and vector respectively.

The time variable t has the values $0,1,\ldots$. A Markov strategy $\pi^{(n)}$ for player n is a sequence of policies: $\pi^{(n)} = (\mu_0^{(n)}, \mu_1^{(n)}, \ldots)$, $\forall_{t\geq 0} [\mu_t^{(n)} \in M^{(n)}]$; the set consisting of all strategies of this kind we call $R^{(n)}$.
A (simultaneous) Markov strategy $\underline{\pi} \in R := \overset{\infty}{\underset{n=1}{\times}} R^{(n)}$ is represented by a sequence made up of the strategy for player 1, the strategy for player 2, and so forth: $\underline{\pi} = (\pi^{(1)}, \pi^{(2)}, \ldots)$. We can also write this as $\underline{\pi} = (\underline{\mu}_0, \underline{\mu}_1, \ldots)$: a sequence of simultaneous policies beginning at time $t = 0$; obviously $(\pi^{(n)})_t = \mu_t^{(n)}$. Note: when speaking about $\underline{\pi} \in R$, we call its time components explicitly $\underline{\mu}_0$, $\underline{\mu}_1$ and so on.

By choosing a strategy $\underline{\pi} \in R$ and a starting state $i \in S$ we determine a stochastic process X_t ($t = 0,1,\ldots$) on S, in the same way as in 2.1: here we have $P(\underline{\mu}_t)$ instead of $P(f_t,g_t)$. The probability measure for this process will be denoted by $P_i^{\underline{\pi}}$; $E_i^{\underline{\pi}}$ is the corresponding expectation operator.

From now on it is assumed that the functions r_n have the charge structure:

Assumption A: $\forall_{n\in\mathbb{N}} \forall_{i\in S} \forall_{\underline{\pi}\in R} [w_n(i,\underline{\pi}) := E_i^{\underline{\pi}} \sum_{t=0}^{\infty} |r_n(X_t,\underline{\mu}_t)| < \infty]$.

Some more definitions:

$$\forall_{n\in\mathbb{N}} \forall_{i\in S} \forall_{\underline{\pi}\in R} [v_n(i,\underline{\pi}) := E_i^{\underline{\pi}} \sum_{t=0}^{\infty} r_n(X_t,\underline{\mu}_t)] ;$$

$$\forall_n \forall_{\pi=(\mu_0,\mu_1,\ldots)\in R^{(n)}} \forall_t [\pi(t) := (\mu_t,\mu_{t+1},\ldots)] ;$$

$$\forall_{\underline{\pi}=(\underline{\mu}_0,\underline{\mu}_1,\ldots)\in R} \forall_t [\underline{\pi}(t) := (\underline{\mu}_t,\underline{\mu}_{t+1},\ldots)] ;$$

the replacement of x_n by x' in a function $h(\underline{x}) = h(x_1,x_2,\ldots)$ is denoted by $h(\underline{x};n:x')$; we use the same notation in $P_i^{\cdot}$ and $E_i^{\cdot}$;

$$\forall_{\underline{\pi}} \forall_t [S^{\underline{\pi}}(t) := \{j \in S \mid \exists_{i\in S} [P_i^{\underline{\pi}}(X_t = j) > 0]\}]$$

$$\forall_n \forall_{\underline{\pi}} \forall_t [S_n^{\underline{\pi}}(t) := \{j \in S \mid \exists_{i\in S} \exists_{\pi'\in R^{(n)}} [P_i^{(\underline{\pi};n:\pi')}(X_t = j) > 0]\}] ;$$

$$\forall_t [S(t) := \{j \in S \mid \exists_{i\in S} \exists_{\underline{\pi}'\in R} [P_i^{\underline{\pi}'}(X_t = j) > 0]\}] .$$

2. Characterization of v-semi-persistent and v-subgame perfect equilibrium points

A strategy $\underline{\pi} \in R$ is called a (Nash) equilibrium point iff

$$\forall_{n\in\mathbb{N}} \forall_{j\in S} \forall_{\pi'\in R^{(n)}} [v_n(j,\underline{\pi}) \geq v_n(j,\underline{\pi};n:\pi')].$$

We set about in the same way as in 2.3.
In this subsection the functions $v_n: S \times (\mathbb{N} \cup \{0\}) \to \mathbb{R}$, $n \in \mathbb{N}$, are supposed to be given unless they are specified; they must satisfy

Assumption B: $\forall_n \forall_t \forall_{\underline{\mu}\in M} [P(\underline{\mu})|v_n(t)| < \infty$ (componentwise)]; we define $\underline{v} := (v_1,v_2,\ldots)$.

By taking $\forall_{n\in\mathbb{N}}\,\forall_{j\in S}\,\forall_{t\geq 0}\,[v_n(j,t) := v_n(j,\underline{\pi}(t))]$ for some $\underline{\pi} \in R$ we can satisfy B: for now $|v_n(i,t)| \leq w_n(i,\underline{\pi}(t))$, so

$$(P(\underline{\mu})|v_n(t)|)(j) \leq (P(\underline{\mu})w_n(\underline{\pi}(t)))(j) + |r_n(j,\underline{\mu})| = w_n(j,\underline{\mu}\underline{\pi}(t)) < \infty .$$

$\underline{\pi} \in R$ is called v-semi-persistent iff

$$(3.2.1)\quad \forall_{n\in\mathbb{N}}\,\forall_{t\geq 0}\,\forall_{j\in S_n^{\underline{\pi}}(t)}\,\forall_{\pi'\in R^{(n)}}\,[v_n(j,\underline{\pi}(t)) = v_n(j,t) \geq v_n(j,\underline{\pi}(t);n:\pi')];$$

$\underline{\pi} \in R$ is called v-saddling iff

$$(3.2.2)\quad \forall_n\,\forall_t\,\forall_{j\in S_n^{\underline{\pi}}(t)}\,\forall_{\mu\in M^{(n)}}\,[r_n(j,\underline{\mu}_t) + (P(\underline{\mu}_t)v_n(t+1))(j) = v_n(j,t) \geq$$

$$\geq r_n(j,\underline{\mu}_t;n:\mu) + (P(\underline{\mu}_t;n:\mu)v_n(t+1))(j)] ;$$

$\underline{\pi} \in R$ is called v-asymptotically definite iff

$$(3.2.3)\quad \forall_n\,\forall_t\,\forall_{j\in S_n^{\underline{\pi}}(t)}\,\forall_{\pi'\in R^{(n)}}\,[\lim_{k\to\infty} E_j^{\underline{\pi}(t)} v_n(X_k,t+k) = 0 =$$

$$= \lim_{k\to\infty} E_j^{(\underline{\pi}(t);n:\pi')} v_n^-(X_k,t+k)] .$$

As the analogue of theorem (2.3.4) we can now give the characterization of v-semi-persistent strategies, with entirely similar proof.

(3.2.4) Theorem. If $\underline{\pi} \in R$, then

$\underline{\pi}$ is v-semi-persistent $\leftrightarrow$ $\underline{\pi}$ is v-saddling and v-asymptotically definite.

Proof. Assume $n\in\mathbb{N}$, $t \geq 0$ and $j \in S_n^{\underline{\pi}}(t)$ arbitrarily chosen.
$\Rightarrow$: i) (*v-saddling*). Let $\mu \in M^{(n)}$; (3.2.1) left side gives (with reasoning similar to the one in (2.3.4))

$$v_n(j,t) = v_n(j,\underline{\pi}(t)) = r_n(j,\underline{\mu}_t) + (P(\underline{\mu}_t)v_n(\underline{\pi}(t+1)))(j) =$$

$$= r_n(j,\underline{\mu}_t) + (P(\underline{\mu}_t)v_n(t+1))(j) ;$$

and

$$v_n(j,t) \geq v_n(j,\underline{\pi}(t);n:\mu\pi^{(n)}(t+1)) =$$

$$= r_n(j,\underline{\mu}_t;n:\mu) + (P(\underline{\mu}_t;n:\mu)v_n(\underline{\pi}(t+1)))(j) =$$

$$= r_n(j,\underline{\mu}_t;n:\mu) + (P(\underline{\mu}_t;n:\mu)v_n(t+1))(j) .$$

ii) (*v-asymptotic definiteness*). Choose $\pi' = (\mu_0', \mu_1', \ldots) \in R^{(n)}$. We have

$$\left|E_j^{\underline{\pi}(t)} v_n(X_k, t+k)\right| \leq E_j^{\underline{\pi}(t)} \left|v_n(X_k, \underline{\pi}(t+k))\right| \leq$$

$$\leq E_j^{\underline{\pi}(t)} \sum_{m=k}^{\infty} \left|r_n(X_m, \underline{\mu}_{t+m})\right| \to 0 \quad (k \to \infty)$$

according to A; and (remember $a \geq b \Rightarrow a^- \leq |b|$)

$$0 \leq E_j^{(\underline{\pi}(t);n:\pi')} v_n^-(X_k, t+k) \leq E_j^{(\underline{\pi}(t);n:\pi')} \left|v_n(X_k, \underline{\pi}(t+k);n:\pi'(k))\right|$$

$$\leq E_j^{(\underline{\pi}(t);n:\pi')} \sum_{m=k}^{\infty} \left|r_n(X_m, \underline{\mu}_{t+m};n:\mu_m')\right| \to 0 \quad (k \to \infty) .$$

$\Leftarrow$: Let $\pi' = (\mu_0', \mu_1', \ldots) \in R^{(n)}$. Now

$$v_n(j,t) = r_n(j, \underline{\mu}_t) + (P(\underline{\mu}_t) v_n(t+1))(j) = \ldots = v_n(j, \underline{\pi}(t)) +$$

$$+ \lim_{k\to\infty} E_j^{\underline{\pi}(t)} v_n(X_k, t+k) = v_n(j, \underline{\pi}(t)) ;$$

and

$$v_n(j,t) \geq r_n(j, \underline{\mu}_t;n:\mu_0') + (P(\underline{\mu}_t;n:\mu_0') v_n(t+1))(j) \geq$$

$$\geq \ldots \geq v_n(j, \underline{\pi}(t);n:\pi') - \lim_{k\to\infty} E_j^{(\underline{\pi}(t);n:\pi')} v_n^-(X_k, t+k) = v_n(j, \underline{\pi}(t);n:\pi').$$

Since $n \in \mathbb{N}$, $t \geq 0$ and $j \in S_n^{\underline{\pi}}(t)$ were arbitrary, the proof is complete. □

If $\underline{\pi} \in R$ satisfies (3.2.1), (3.2.2) or (3.2.3) with $S(t)$ replacing $S_n^{\underline{\pi}}(t)$, we call $\underline{\pi}$ <u>v-subgame perfect</u>, <u>v-overall saddling</u> or <u>v-overall asymptotically definite</u> respectively.
We can now present the characterization of <u>v</u>-subgame perfect:

(3.2.5) <u>Theorem</u>. If $\underline{\pi} \in R$, then $\underline{\pi}$ is <u>v</u>-subgame perfect iff $\underline{\pi}$ is <u>v</u>-overall saddling and <u>v</u>-overall asymptotically definite.

<u>Proof</u>. The assertion is easily checked by replacing $S_n^{\underline{\pi}}(t)$ by $S(t)$ in the proof of (3.2.4). □

3. N-person Markov games; Markov decision processes

At first we consider Markov games with N players, $N \in \mathbb{N}$, and assume $K_i^{(n)}$ at most countable but nonempty ($1 \leq n \leq N$).
In our model (3.1) we have the following simplifications: we may presume that, if $n > N$, $\forall_{i \in S}\,[K_i^{(n)} = \{0\}]$, and henceforth abstain from considering that case; in general we can take $\Gamma_i^{(n)} = P(K_i^{(n)})$ (power set), and $M_i^{(n)}$ as the set of randomized actions for player n and state $i \in S$, $1 \leq n \leq N$; we define $F^{(n)} := M^{(n)}$ so as to make clear the analogy with F and G from subsection 2.1. For all $\underline{f} = (f_1,\ldots,f_N) \in F^{(1)} \times\ldots\times F^{(N)} =: F$ it holds that

$$(P(\underline{f}))_{ij} = \sum_{(k_1,\ldots,k_N)} p(i,j,k_1,\ldots,k_N) f_1(i,k_1)\ldots f_N(i,k_N) ,$$

$$r_n(i,\underline{f}) = \sum_{(k_1,\ldots,k_N)} r_n(i,k_1,\ldots,k_N) f_1(i,k_1)\ldots f_N(i,k_N) .$$

We can now apply the theory of 3.2 to N-person Markov games. Without objection $N = 1$ can be substituted: in that case we are concerned with one-person Markov games, that is: *Markov decision processes*. Given are now S, K_i countable ($i \in S$), probabilities $p(i,j,k)$ ($i,j \in S$, $k \in K_i$) and rewards $r(i,k)$ ($i \in S$, $k \in K_i$); for all $f \in F$ we have $P(f)$ and $r(f)$ as above.
A Markov strategy $\pi \in R$ may be written as $\pi = (f_0,f_1,\ldots)$, $\forall_t\,[f_t \in F]$. Assumption A reads here as follows:

$$\forall_{\pi \in R}\,\forall_{i \in S}\,[w(i,\pi) := E_i^{\pi} \sum_{t=0}^{\infty} |r(X_t,f_t)| < \infty] ,$$

so that $v(i,\pi) := E_i^{\pi} \sum_{t=0}^{\infty} r(X_t,f_t)$ exists for all $\pi \in R$ and $i \in S$; finally we have $\forall_{\pi}\,\forall_t\,[S_1^{\pi}(t) = S(t)]$: this means that the concepts v-semi-persistent and v-subgame perfect coincide.

Take $v(j,t) := \sup_{\pi \in R} v(j,\pi) =: v(j)$, then $|v| < \infty$ as well as $\forall_{f \in F}\,[P(f)|v| < \infty]$, according to [2] lemma 2 (this may be seen by taking in the model of [2] $\forall_i\,[L(i) := \{0\}]$).
Applying (3.2.4) (= (3.2.5)) to Markov decision processes we get (the right-hand parts of the assertions can be omitted!)

(3.3.1) Theorem. If $\pi \in R$, then

$$\forall_t \forall_{j\in S(t)} [v(j,\pi(t)) = v(j)] \Leftrightarrow \begin{cases} \forall_t \forall_{j\in S(t)} [r(j,f_t) + (P(f_t)v)(j) = v(j)] \\ \forall_t \forall_{j\in S(t)} [\lim_{k\to\infty} E_j^{\pi(t)} v(X_k) = 0] . \end{cases}$$

Compare (3.3.1) with [3] Theorem 1 ("π optimal iff π is value-conserving and equalizing"): there instead of A merely the charge structure property for *non*randomized strategies is assumed; Theorem 1 provides the analogue of (3.3.1) with $S^{\pi}(t)$ replacing $S(t)$, for strategies of that kind.

Acknowledgement

The author thanks mr. L.P.J. Groenewegen and Dr. J. Wessels, under whose supervision this research was made.

References.

[1] L.P.J. Groenewegen, Markov games: properties of and conditions for optimal strategies. Memorandum COSOR 76-24, Eindhoven University of Technology, Department of Mathematics, November 1976.

[2] L.P.J. Groenewegen and J. Wessels, On the relation between optimality and saddle-conservation in Markov games. Memorandum COSOR 76-14, Eindhoven University of Technology, Dept. of Math., October 1976.

[3] L.P.J. Groenewegen, Convergence results related to the equalizing property in a Markov decision process. Memorandum COSOR 75-18, Eindhoven University of Technology, Department of Mathematics, October 1975.

[4] J.C.W.H.M. Pulskens, Stochastische niet-coöperatieve twee personen spelen (Master's thesis, in Dutch), Eindhoven University of Technology, Department of Mathematics, August 1976.

[5] R. Selten, Reexamination of the perfectness concept for equilibrium points in extensive games. Int. J. Game Theory 4 (1975), 25-55.