Markov Decision Evolutionary Games

Eitan Altman and Yezekael Hayel

March 28, 2008

Abstract

We present a class of evolutionary games involving large populations that have many pairwise interactions between randomly selected players. The fitness of a player depends not only on the actions chosen in the interaction but also on the individual state of the players. Players have finite life time and take during which they participate in several local interactions. The actions taken by a player determine not only the immediate fitness but also the transition probabilities to its next individual state. We define and characterize the Evolutionary Stable Strategies (ESS) for these games and propose a method to compute them. We illustrate the model and results through a networking problem.

1 Introduction

Evolutionary games have been developed by J. Meynard Smith to model the evolution of population sizes as a result of competition between them that occurs through many local pairwise interactions, i.e. interactions between randomly chosen pairs of individuals. Central in evolutionary games is the concept of Evolutionary Stable Strategy, which is a distribution of (deterministic or mixed) actions such that if used, the population is immune against penetration of mutations. This notion is stronger than that of Nash equilibrium as ESS is robust against a deviation of a whole fraction of the population where as the Nash equilibrium is defined with respect to possible deviations of a single player. A second foundation of evolutionary games is the replicator dynamics that describes the dynamics of the sizes of the populations as a result of the fitness they receive in interactions. Meynard Smith formally introduced both, without needing an explicit modeling of stochastic features. We shall call this the deterministic evolutionary game.

Randomness is implicitly hinted in the requirement of robustness against mutations, and indeed the ESS is defined through robustness against any mutation. Random aspects can be explicitly included in the modeling of evolutionary games. We first note that since deterministic evolutionary games deal with large populations, they may provide an interpretation of the deterministic game as a limit smaller games that included randomness that has been averaged out by some strong law of large numbers. Such an interpretation can be found in [9].

Yet, other sources of randomness have been introduced into evolutionary games. Some authors have added small noise to the replicator dynamics in order to avoid the problem of having the dynamics stuck in some local minimum, see [13, 10, 11] and references therein. The ESS can then be replaced by other notions such as the the SSE [10].

In this paper we introduce another class of stochastic evolutionary games, which we call "Markov Decision Evolutionary Games" (MDEG). There are again many local interactions among individuals belonging to large populations of players. Each individual has a finite life time during which (i) it may move among different individual states, and (ii) it interacts several times with other users; the actions of the player along with those with which it interacts determine not only the immediate fitness of the player but also the transition probabilities to the next state it will have. An individual can be seen as trying to maximize the expected sum of its immediate fitness during his life time.

A simple application of an MDEG to mobile communications has been introduced in [4]. Mobile terminals transmit packets occasionally. Their destination occasionally may receive simultaneously a transmission from another terminal which results in a collision. It is assumed however that even when packets collide, one of the packets can be received correctly if transmitted at a higher power. The immediate fitness rewards successful transmissions and penalizes energy consumption. Each mobile decides at each slot what its power level will be. In [4], this decision is allowed to depend on the depletion level of the battery, which serves as the "individual state". The battery is considered to be either in the state "Full" (F) in which case there are two power levels available, or "Almost Empty" (AE) in which only the weak power level is available. Transmission at high power at state F results in a larger probability of moving to state AE. We present in Subsection 3.4 an extension of this problem.

In contrast to the above simple application, in which decisions can be taken only at one individual state, we develop in this paper a general theory for computing an ESS where decisions can be taken at all individual states, where the fitness depends not only on the actions but also on the individual states of the interacting individuals, and where the available actions may depend on the individual state.

An interesting application of MDEG is the repeated game version of the well known Hawk and Dove game in which some of the features of MDEG are already present, [5, 6, 7].

The structure of the paper is as follows. After a brief description of standard evolutionary games presented next, we introduce in Section 3 the model of MDEG and define the main solution concept, the Weak ESS, in that context. We provide a method for computing this ESS in Section 4 which provide ESS within mixed strategies. Some alternative view point through dynamic programming is introduced in Section 11 to provide a characterization of the ESS in stationary policies. The definition and characterization of Strong ESS are provided in Section 6. In Section 7 we compare the mixed and the stationary ESS in an energy control problem. This is followed by a concluding section.

2 Reminder on (standard) Evolutionary Games (EG)

Consider a large population of players. Each individual needs occasionally to take some action. We focus on some (arbitrary) tagged individual. Occasionally, the action of some N (possibly random number of) other individuals interact with the action of that individual. We define by $J(p,q)$ the expected payoff for our tagged individual if it uses a strategy (also called policy) p when meeting another individual who adopts the strategy q. This payoff is called "fitness" and strategies with larger fitness are expected to propagate faster in a population. p and q belong to a set K of available strategies. In the standard framework for evolutionary games there are a finite number of so called "pure strategies", and a general strategy of an individual is a probability distribution over the pure strategies. An equivalent interpretation of strategies is obtained by assuming that individuals choose pure strategies and then the probability distribution represents the fraction of individuals in the population that choose each strategy. Note that J is linear in p and q.

Suppose that the whole population uses a strategy q and that a small fraction ϵ (called "mutations") adopts another strategy p. Evolutionary forces are expected to select against p if

$$J(q, \epsilon p + (1-\epsilon)q) > J(p, \epsilon p + (1-\epsilon)q) \tag{1}$$

Definition 2.1 *A strategy q is said to be ESS if for every $p \neq q$ there exists some $\bar{\epsilon}_y > 0$ such that (1) holds for all $\epsilon \in (0, \bar{\epsilon}_y)$.*

In fact, we expect that if

$$\text{for all } p \neq q, \quad J(q,q) > J(p,q) \tag{2}$$

then the mutations fraction in the population will tend to decrease (as it has a lower reward, meaning a lower growth rate). q is then immune to mutations. If it does not but if still the following holds,

$$\text{for all } p \neq q, \quad J(q,q) = J(p,q) \text{ and } J(q,p) > J(p,p) \tag{3}$$

then a population using q are "weakly" immune against a mutation using p since if the mutant's population grows, then we shall frequently have individuals with strategy q competing with mutants; in such cases, the condition $J(q,p) > J(p,p)$ ensures that the growth rate of the original population exceeds that of the mutants. We shall need the following characterization:

Theorem 2.1 [1, Proposition 2.1] or [2, Theorem 6.4.1, page 63] *A strategy q is ESS if and only if it satisfies (2) or (3).*

Corollary 2.1 *(2) is a sufficient condition for q to be an ESS. A necessary condition for it to be an ESS is*

$$\textit{for all } p \neq q, \quad J(q,q) \geq J(p,q) \tag{4}$$

The conditions on ESS can be related and interpreted in terms of a Nash equilibrium in a matrix game. The situation in which an individual, say player 1, is faced with a member of a population in which a fraction p chooses strategy A is then translated to playing the matrix game against a second player who uses mixed strategies (randomizes) with probabilities p and $1-p$, resp.

3 Model

We use a hierarchical description of the system composed of a model for the individual player and a global model for aggregating individual's behavior.

3.1 Model for Individual player

A player arrives at some random time t_0. It has a clock that is responsible to the times at which interactions with other players occur. It is involved in interactions that occur according to a Poisson process with rate λ. After a random number of clicks, the player leaves the system and is replaced be another one. This will be made precise below. During the player's life time, each time the timer clicks, the player interacts with another randomly selected player.

We associate with each player a Markov Decision Process (MDP) embedded at the instants of the clicks.

The parameters of the MDP are given by the tuple $\{\mathcal{S}, \mathcal{A}, Q\}$ where

- $\mathcal{S}$ is the set of possible individual states of the player
- $\mathcal{A}$ is the set of available actions. For each state s, a subset $\mathcal{A}_s$ of actions is available.
- Q is the set of transition probabilities; for each $s, s' \in \mathcal{S}$ and $a \in \mathcal{A}_s$, $Q_{s'}(s,a)$ is the probability to move from state s to state s' taking action a. $\sum_{s' \in S} Q_{s'}(s,a)$ is allowed to be smaller than 1.

Define further

- The set of policies is $\mathcal{U}$. A general policy u is a sequence $u = (u_1, u_2, \ldots)$ where u_i is a distribution over action space $\mathcal{A}$ at time i. The dependence on time is a local one: it concerns only the individual's clock; a player is not assumed to use policies that make use of some global clocks.
- The subset of mixed (resp. pure or deterministic) policies is $\mathcal{U}_M$ (resp. $\mathcal{U}_D$). We define also the set of stationary policies $\mathcal{U}_S$ where such policy does not depend on time.

Occupation measure Often we encounter the notion of individual states in evolutionary games; but usually the population size at a particular state is fixed. In our case the choices of actions of an individual determine the fraction of time it would spend at each state. Hence the fraction of the whole population

that will be at a given state may depend on the distribution of strategies in the population. In order to model this dependence we first need to describe the expected amount of time $f_{\eta,u}(s)$ that an individual spends at a given state s when it follows a strategy u and its initial state at time 1 is distributed according to a probability η over $\mathcal{S}$. More generally, we define $f_{\eta,u}(s,a)$ the expected number of time units during which it is at state s and it chooses action a. $f_u := \{f_{\eta,u}(s,a)\}$ is called the occupation measure corresponding to a policy u.

Define $p_t(\eta,u;s,a) = I\!P_{\eta,u}(X_t = s, A_s = a)$ the probability for a user to be in state s, at time t, using action a under policy u when the initial state has a probability distribution η. Further define $p_t(\eta,u;s) = \sum_a p_t(\eta,u;s,a)$. $p_t(\eta,u;s)$ is a sub-probability measure as $\sum_{s\in\mathcal{S}} p_t(\eta,u;s)$ may be smaller than one. The lifetime of an individual is identified as the time interval before X_t leaves $\mathcal{S}$. We have

$$f_{\eta,u}(s,a) = \sum_{t=1}^{+\infty} p_t(\eta,u;s,a), \quad f_{\eta,u}(s) = \sum_{t=1}^{+\infty} p_t(\eta,u;s).$$

Define the expected lifetime of a player corresponding to a given η and u as $T_{\eta,u} = \sum_s f_{\eta,u}(s)$. We shall assume throughout that for a given η, $\sup_{u\in U} T_{\eta,u}$ is finite. We know from [12] that $\sup_{u\in U} T_{\eta,u} = \max_{u\in U_D} T_{\eta,u}$ so the assumption is equivalent to requesting:

A1: $T_{\eta,u}$ is finite for all $u \in U_D$.

3.2 Interactions and System model

We have a large population of individuals. As in standard evolutionary games, there are many pairwise interactions between randomly selected pairs.

Let $r(s,a,s',b)$ be the immediate reward that a player receives when it is at state s and it uses action a while interacting with a player who is in state s' that uses action b.

Denote by $\alpha(u) = \{\alpha(u;s,a)\}$ the system state: $\alpha(u;s,a)$ is the fraction of the population at individual state s and that use action a when all the population uses strategy u. We shall add the index t to indicate a possible dependence on some time.

Consider an arbitrary tagged player and let S_t and A_t be its state and action at time t (as measured on its individual clock). Then his expected immediate reward at that time is given by

$$R_t = \sum_{s,a} \alpha_t(u;s,a) r(S_t, A_t, s, a)$$

Assume now that a player arrives at the system at time 1. The global expected fitness when using a policy v is then

$$F_\eta(v,u) = \sum_{t=1}^{\infty} E_{\eta,v}[R_t].$$

When η is concentrated on state s we write with some abuse of notation $F_s(v,u) = F_\eta(v,u)$. We shall often omit the index η (in case it is taken to be fixed).

Unless stated differently, we shall make throughout the following assumption:

A2: When the whole population uses a policy u, then at any time t which is either fixed or is an individual time of an arbitrary player, $\alpha_t(u)$ is independent of t and is given by

$$\alpha_t(u;s,a) = \frac{f_{\eta,u}(s,a)}{T_{\eta,u}}$$

for all s, a.

Under this assumption, the global expected fitness simplifies to

$$F(v,u) = \sum_{t=1}^{\infty} E_{\eta,v} R_t = \sum_{s,a} f_{\eta,v}(s,a) \frac{1}{T_{\eta,u}} \sum_{s',a'} f_{\eta,u}(s',a') r(s,a,s',a') \qquad (5)$$

Assumption A2 would not hold if the policy of a player could depend on the absolute time or on the behavior (i.e. the actions) of other players. For example, in the standard replicator dynamics, the policy of a player adapts to the instantaneous fitness which depends also on the actions of the other players in the population. Thus A2 does not hold there.

On the other hand, since players of a given class are undistinguishable, and since the lifetime distribution of a mobile depends only on his local time, we may expect Assumption A2 to hold. Checking A2 is beyond the scope of the paper.

Definition 3.1 *We shall say that two strategies u and u' are equivalent if the corresponding occupation measures are equal. We shall write $u =_e u'$.*

Note that if u and u' are equivalent policies for a given player then for any v used by the rest of the population, the fitness under u and under u' are the same.

3.3 Defining the Weak ESS

With the expression (5) for the fitness, we observe that we are again in the framework of Section 2 and can use Definition of Theorem 2.1 for the weak ESS in the MDEG:

Definition 3.2 *A strategy u is a weak ESS, denoted by WESS, for the MDEG if and only if it satisfies one of the following:*

$$\text{for all } v \neq_e u, \quad F(u,u) > F(u,v) \qquad (6)$$

$$\text{for all } v \neq_e u, \quad F(u,u) = F(v,u) \text{ and } F(u,v) > F(v,v) \qquad (7)$$

This notion of weakness stable strategy is related to the equivalent class in terms of occupation measure.

Lemma 3.1 *A policy u is WESS for the MDEG if and only if it is an WESS for the following modified global fitness function:*

$$\widetilde{F}(v,u) = \sum_{s,a} f_{\eta,v}(s,a) \sum_{s',a'} f_{\eta,u}(s',a') r(s,a,s',a') \tag{8}$$

The advantage of the new form of the fitness function (8) is that it is bilinear in the occupation measures of the players that interact with each other. The set of occupation measures will be shown to be a polytope whose extreme points correspond to strategies in U_D. This will allow us to transform the MDEG to a standard EG.

We could use the following as an equivalent Definition of WESS for MDEG.

Theorem 3.1 *A strategy u is said to be WESS if for every $v \neq_e u$ there exists some $\overline{\epsilon}_v > 0$ such that the following holds for all $\epsilon \in (0, \overline{\epsilon}_v)$:*

$$F(u, \epsilon u + (1-\epsilon)v) > F(u, \epsilon u + (1-\epsilon)v) \tag{9}$$

In (9) we use a convex combination of two policies. We delay the definition of this to the next section (see Remark 4.1).

3.4 Application to Energy Control in Wireless Networks

We next illustrate the MDEG setting with a problem that arises in dynamic power control in mobile networks. A special case of this framework (where a choice between several control actions exists in one state only) has been studied in [4].

Users participate in local competitions for the access to a shared medium in order to transmit their packets. An individual state of each mobile represents the energy level at the user's battery which, for simplicity, we assume to take finitely many values, denoted by $\mathcal{S} = \{0, \ldots, S\}$. In each state $s \in \mathcal{S} \setminus \{0\}$, each mobile has two available actions h and l which correspond respectively to high power p_H and low power p_L. We consider an Aloha-type game where a mobile transmits a packet with success during a slot if:

- the mobile is the only one to transmit during this slot, with probability p
- the mobile transmits with high power and all others transmitting mobiles use low power.

The reward function r depends only on the transmission powers, that is, the action of the mobile and the one in competition with him. Then we have:

$$r(s,a,s',a') = p + (1-p)\mathbb{1}_{((a=h) \quad \text{and} \quad (a'=l))}.$$

For each state $s \in \mathcal{S} \setminus \{0\}$, the transition probability $Q_{s'}(s,a)$ may be non-zero (for both $a \in \{l,h\}$) only for $s' \in \{s, s-1\}$. Then, as the two possible transitions are to remain at the same energy level or move to the next lower one, we simplify the notation and use $Q(s,a)$ to denote the probability of remaining at energy level s using action a. We have the following assumptions on the transition probabilities which are motivated by the application.

- For all state $s \in \mathcal{S} \setminus \{0\}$, we have $Q(s,h) < Q(s,l)$ because using less power induces higher probability to remain in the same energy level.
- For all state $s \in \mathcal{S} \setminus \{0\}$ and for both actions $a \in \{l,h\}$, we have $Q(s,a) > Q(s-1,a)$ because less battery energy the mobile has, less is the probability to remain at the same energy level.

We consider the three states $S = \{0,1,2\}$ and then the set U_D of the deterministic policies is composed of the following four couples : (l,l), (l,h), (h,l) and (h,h); where the first element is the action taken in state 1 and the second one is the action taken by a mobile in state 2. We denote these strategies by u_1, u_2, u_3 and u_4. We recall that in state 0, there is no more actions available as the mobile has no more energy.

We present the solution of this problem in Subsection 4.2.

4 Computing the WESS

Define the set of occupation measures achieved by all (individual) policies in some subset $U' \subset U$ as

$$\mathcal{L}_\eta(U') = \bigcup_{u \in U'} f_{\eta,u}(s,a).$$

It will turn out that the expected fitness of an individual (defined in next subsection) will depend on the strategy u of that individual only through $f_{\eta,u}$. We are therefore interested in the following characteristic of $\mathcal{L}_\eta(U)$ (see [12, 8]):

Lemma 4.1 *$\mathcal{L}_\eta(U)$ equals to the set Q_η defined as the set of $\alpha = \{\alpha(s,a)\}$ satisfying*

$$\sum_{s' \in S} \sum_{a \in A'_s} \alpha(s',a)[\delta_{s'}(s) - Q_{s'}(s,a)] = \eta(s), \forall s, \quad \alpha(s,a) \geq 0, \forall s,a. \tag{10}$$

where $\delta_{s'}(s)$ is the dirac distribution in state s'.
(ii) We have: $\mathcal{L}_\eta(U) = \mathcal{L}_\eta(U_S) = co\mathcal{L}_\eta(U_D)$ where $co\mathcal{L}_\eta(U_D)$ is the convex hull of $\mathcal{L}_\eta(U_D)$.
(iii) For any $\alpha \in \mathcal{L}_\eta(U)$, define the individual stationary policy $u \in \mathcal{U}_S$ by

$$u_s(a) = \frac{\alpha(s,a)}{\sum_{a \in A_s} \alpha(s,a)}.$$

Then $f_{\eta,u} = \alpha$.

4.1 Transforming the MDEG into a standard EG

Consider the following standard evolutionary game **EG**:

- the finite set of actions of a player is U_D,
- the fitness of a player that uses $v \in U_D$ when the other use a policy $u \in U_S$ is given by (8).
- Enumerate the strategies in U_D such that $U_D = (u_1, ..., u_m)$.
- Define $\gamma = (\gamma_1, ..., \gamma_m)$ where γ_i is the fraction of the population that uses u_i. γ can be interpreted as a mixed strategy which we denote by $\widehat{\gamma}$.

Remark 4.1 *Here the convex combination $\epsilon\widehat{\gamma}+(1-\epsilon)\widehat{\gamma}'$ of the two mixed strategies $\widehat{\gamma}$ and $\widehat{\gamma}'$ is simply the mixed strategy whose ith component is given by $\epsilon\gamma_i + (1-\epsilon)\gamma_i'$, $i = 1, ..., m$.*

Combining Lemma 4.1 with Lemma 3.1 we obtain:

Proposition 4.1 *Let $\widehat{\gamma}$ be an ESS for the game* **EG**. *Then it is a WESS for the original MDEG.*

4.2 Application to Energy Control in Wireless Networks (continued)

We pursue the example described in Section 3.4 applying the latest proposition in order to obtain the WESS for this MDEG. Indeed, we will find the WESS for the related EG game which will be written as a matrix game with dimension 4. In order to find the equilibrium of this matric game, we have to compute the fitness $\widetilde{F}(u, v)$ for all policies u and v.

In a first step, we have to compute the occupation measure f_u corresponding to each policy $u \in \{u_1, u_2, u_3, u_4\}$; for that we need the probability for a user to be in each state, at time t, using action a under policy u. At initial time $t = 0$, a mobile always starts with a battery full of energy, that is $\eta = (0, 0, 1)$. We describe the matrix game with the two following matrices:

$$\widetilde{F_1} = \begin{pmatrix} (X_1+X_3)^2 & (X_1+X_3)(X_1+X_4) & (X_1+X_3)(X_2+X_3) & (X_1+X_3)(X_2+X_4) \\ (X_1+X_4)(X_1+X_3) & (X_1+X_4)^2 & (X_1+X_4)(X_2+X_3) & (X_1+X_4)(X_2+X_4) \\ (X_2+X_3)(X_1+X_3) & (X_2+X_3)(X_1+X_4) & (X_2+X_3)^2 & (X_2+X_3)(X_2+X_4) \\ (X_2+X_4)(X_1+X_3) & (X_2+X_4)(X_1+X_4) & (X_2+X_4)(X_2+X_3) & (X_2+X_4)^2 \end{pmatrix}$$

and

$$\widetilde{F_2} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ X_4(X_1+X_3) & X_4X_1 & X_4X_3 & 0 \\ X_2(X_1+X_3) & X_2X_1 & X_2X_3 & 0 \\ (X_2+X_4)(X_1+X_3) & (X_2+X_4)X_1 & (X_2+X_4)X_3 & 0 \end{pmatrix}$$

with

$$X_1 = \frac{1}{1-Q(1,l)}, \quad X_2 = \frac{1}{1-Q(1,h)}, \quad X_3 = \frac{1}{1-Q(2,l)}, \quad X_4 = \frac{1}{1-Q(2,h)}.$$

The computation of this matrix is described in appendix A.

Then we obtain the following modified fitnesses depending on the policies in the following matrix :

$$\widetilde{F} = p\widetilde{F_1} + (1-p)\widetilde{F_2}.$$

The WESS of the MDEG which model energy control behaviors in wireless networks is obtained in finding the WESS of the standard EG with the matrix of fitnesses given by $\widetilde{F}$.

5 WESS and Dynamic Programming

In Section 4 we transformed the MDEG into a standard EG whose actions correspond to the set of deterministic policies. Mixed policies are then interpreted as selecting randomly one deterministic policy.

The set of such policies is huge. We are therefore interested in obtaining WESS policies within other classes of policies.

We next consider the class of stationary policies. It follows from Lemma 4.1 (ii) that there is a stationary WESS policy if and only if there is an equivalent mixed WESS policy. We shall restrict in the rest of the section to stationary (possibly randomized) policies.

Theorem 5.1 *(i) A necessary condition for a policy u to be WESS is that $F(u,u) \geq F(v,u)$ for all stationary v.*
(ii) Assume that the following set of dynamic programming equations holds: For all state $s \in \mathcal{S}$,

$$F_s(u,u) = \max_a \left[r(u;s,a) + \sum_{s'} Q_{s'}(s,a) F_{s'}(u,u) \right] . \tag{11}$$

Then $F(u,u) \geq F(v,u)$ for any v.
(iii) If $\eta(s) > 0$ for all s, then the converse also holds: and (11) is equivalent to $F(u,u) \geq F(v,u)$ for all stationary v.

Proof: Part (i) follows directly from the Definition 3.2 of an WESS for MDEG.
(ii) Now, Eq (11) can be interpreted as a one step of policy iteration performed by a player which solves an MDP given that all other players use u. (11) then states that this player cannot improve further the policy iteration, and therefore $F(u,u) \geq F(v,u)$ for all v, see e.g. [14, p. 72]. Hence (11) implies $F(u,u) \geq F(v,u)$ for all v.

(iii) We show that the converse holds. Assume now that (11) does not hold. Then there is some action a such that for each s there is an action $a = a(s)$ such that

$$F_s(u,u) \leq r(u;s,a) + \sum_{s'} Q_{s'}(s,a) F_{s'}(u,u)$$

with strict inequality in at least one state s. Consider now the policy v that agrees with u from time 2 onwards and that uses action $a(s)$ at state s at time 1. Then it follows that $F_s(u,u) \leq F_s(v,u)$ for all s with strict inequality at least for one s. Since $F_\eta(v,u) = \sum_s \eta(s) F_s(v,u)$ and since $\eta(s)$ are assumed to be strictly positive for all s, this implies that $F_\eta(u,u) < F_\eta(v,u)$. We conclude that indeed (11) is equivalent to $F_\eta(u,u) \geq F_\eta(v,u)$ for all v. $\diamond$

6 Strong ESS

Under our previous definition 3.2 of WESS, a WESS u could be invaded by another policy v provided that u and v have the same occupation measure: $v =_e u$. We now propose a characterization of another version of ESS through dynamic programming for which an ESS u cannot be invaded by any other policy v.

Definition 6.1 *A strategy u is Strong ESS (SESS) for the MDEG if and only if it satisfies one of the following:*

$$\text{for all } v \neq u, \quad F(u,u) > F(v,u) \tag{12}$$

$$\text{for all } v \neq u, \quad F(u,u) = F(v,u) \text{ and } F(u,v) > F(v,v) \tag{13}$$

Theorem 6.1 *Let $\eta(s) > 0$ for all s and assume that (11) holds. Then u is an SESS if one of the following holds.*
(i) For all states s there is a single a attaining the maximum.
(ii) For each policy $v \neq u$ the following holds. For all s

$$F_s(u,v) \leq \max_a \left[r(v;s,a) + \sum_{s'} Q_{s'}(s,a) F_{s'}(v,v) \right] \tag{14}$$

and for some s that may depend on v the inequality is strict.

Proof: Assume that (11) holds. Then $F_s(v,u) \leq F_s(u,u)$ for all s (by the previous Theorem). If $v \neq u$ then there exists some s such that Then

$$F_s(v,u) = r(u;s,v(s)) + \sum_{s'} Q_{s'}(s,a) F_{s'}(v,u) \leq r(u;s,v(s)) + \sum_{s'} Q_{s'}(s,a) F_{s'}(u,u)$$

$$< r(u;s,u(s)) + \sum_{s'} Q_{s'}(s,a) F_{s'}(u,u) = F_s(u,u).$$

The assumption on η implies that $F(v,u) < F(u,u)$. Thus by Theorem 5.1, u is an SESS.
(ii) As in the first part we get $F_s(v,u) \leq F_s(u,u)$. If the inequality is strict for at least one state s then the proof is complete. If not then we have $F_s(v,u) = F_s(u,u)$. From (14) we get similarly $F_s(u,v) \geq F_s(v,v)$ for all s. Since the inequality is strict for at least one state s then together with the condition on η we get $F_s(u,v) > F_s(v,v)$. Thus by (13), u is a SESS. $\diamond$

In [4] we use adynamic programming approach in order to compute the WESS but we limit our study to a stationary policy. In the next section, we compare the result obtain in [4] with the WESS obtained by transforming the MDEG into a standard EG.

7 Comparison

We compare our result obtained in [4] where we used dynamic programming approach and the one described in section 4. The model considered in [4] assumes that the mobile cannot transmit with high power in state 1. Then both model are identical if we restrict the deterministic policies to $u_1 = (l,l)$ and $u_2 = (l,h)$. The WESS of the MDEG is the ESS of a standard EG defined by through the related matrix game (that describes the pairewize interactions):

$$\widetilde{G} = \begin{pmatrix} p(X_1+X_3)^2 & p(X_1+X_3)(X_1+X_4) \\ (X_1+X_3)(p(X_1+X_4)+(1-p)X_4) & p(X_1+X_4)^2+(1-p)X_1X_4 \end{pmatrix}$$

which is the restriction of $\widetilde{F}$ to policies $u_1 = (l,l)$ and $u_2 = (l,h)$. We now look at the ESS $\widehat{\gamma}$ where γ is obtained as the equilibrium of this matrix game. We have two pure ESS: $\gamma = 0$ and $\gamma = 1$. (We use for defining $\gamma = (\gamma_1, \gamma_2)$ just the first parameter γ_1 and denote it with some abuse of notation as γ; we have then $\gamma_2 = 1-\gamma$.)

Proposition 7.1 *We have the following pure SESS:*

- $\gamma = 1$ *if* $p < \frac{X_4}{X_3+\frac{X_4}{X_1}(X_3-X_4)}$ *and only if* $p \leq \frac{X_4}{X_3+\frac{X_4}{X_1}(X_3-X_4)}$,
- $\gamma = 0$ *if* $p > \frac{X_4}{X_3}$ *and only if* $p \geq \frac{X_4}{X_3}$.

Proof: Some straight forward calculation show that the condition $p < \frac{X_4}{X_3+\frac{X_4}{X_1}(X_3-X_4)}$ is equivalent to $F(u_2,u_2) > F(u_1,u_2)$. Hence by Definition (12), u_2 is a SESS if this condition holds, which is equivalent to $\gamma = 1$, and it is not an SESS if the condition with the non-strict inequality fails to hold. The rest is established in the same way. $\diamond$

We now look at the existence and uniqueness of a mixed WESS.

Proposition 7.2 *If* $\frac{X_4}{X_3} < p < \frac{X_4}{X_3+\frac{X_4}{X_1}(X_3-X_4)}$, *the game has the following mixed WESS*

$$\gamma = \frac{-X_1X_4 + pX_1X_3 - X_3X_4 + pX_3^2}{-pX_3X_4 - X_3X_4 + pX_3^2 + pX_4^2}.$$

Proof: The mixed WESS of this game $\widehat{\gamma}$ is obtained, for $\frac{X_4}{X_3+\frac{X_4}{X_1}(X_3-X_4)} < p < \frac{X_4}{X_3}$ by resolving the equation:

$$F(u_1, \widehat{\gamma}) = F(u_2, \widehat{\gamma}),$$

with $F(u_i, \widehat{\gamma}) = (1-\gamma)F(u_i, u_1) + \gamma F(u_i, u_2)$ that is the expected fitness of a player using strategy u_i. This equation has the unique solution

$$\gamma = \frac{-X_1X_4 + pX_1X_3 - X_3X_4 + pX_3^2}{-pX_3X_4 - X_3X_4 + pX_3^2 + pX_4^2}.$$

$\diamond$

We are able to compare the occupation measure in state 2, that is state F in [4]. The ESS computed in [4] is obtained when restricting to stationary policies: each time state F is visited, actions are chosen according to the same probability measure denoted by a parameter β. We denote the policy by β^*. This is different than the WESS obtained here in mixed policies, where there is a single initial randomization between pure strategies, which means that a mobile uses the same action each time state F is reached.

Proposition 7.3 *Policies* β^* *defined in [4] and* $\widehat{\gamma}$ *described in proposition 7.2 are in the same equivalent class,i.e.*

$$\beta^* =_e \widehat{\gamma}.$$

Proof: First, we observe that the stationary WESS β^* corresponds to the pure strategy $\beta = 1$ if and only if

$$\frac{1-Q_F(h)}{2-Q_A-Q_F(h)}(1-p) - \frac{Q_F(l)-Q_F(h)}{1-Q_F(l)}p \geq 0,$$

which, after some basic calculations is equivalent to

$$p \leq \frac{X_4}{X_3+\frac{X_4}{X_1}(X_3-X_4)}.$$

We conclude that if $p \leq \frac{X_4}{X_3+\frac{X_4}{X_1}(X_3-X_4)}$ then $\beta = \gamma = 1$.
Second, we have from [4] that the stationary WESS β^* is the pure strategy corresponding to $\beta = 0$ if and only if

$$p(1-Q_F(h)) - (1-Q_F(l)) \geq 0,$$

which is equivalent to

$$p \geq \frac{X_4}{X_3}.$$

Then if $p \geq \frac{X_4}{X_3}$ we have that $\beta = \widehat{=} 0$. Thus if $p \geq \frac{X_4}{X_3 + \frac{X_4}{X_1}(X_3 - X_4)}$ or $p \geq \frac{X_4}{X_3}$ we have that $\beta = \gamma$ then both policies have the same occupation measure.
We now look at the case when $\frac{X_4}{X_3 + \frac{X_4}{X_1}(X_3 - X_4)} < p < \frac{X_4}{X_3}$. The ESS obtained in [4] is

$$\beta = \frac{(2 - Q_A - Q_F(l))[1 - Q_F(l) - p(1 - Q_F(h))]}{(1 - Q_A)(1 - p)(1 - Q_F(l)) - (Q_F(l) - Q_F(h))((1 - Q_F(l)) - p(1 - Q_F(h)))},$$

which is equal to:

$$\beta = \frac{(X_1 + X_3)(X_4 - pX_3)X_4}{(1 - p)X_3X_4^2 - X_1(X_3 - X_4)(X_4 - pX_3)}.$$

The occupation measure in state F is:

$$T_{\beta^*}(F) = \frac{1}{1 - \beta Q_F(h) - (1 - \beta)Q_F(h)} = \frac{1}{1 - \beta\frac{X_4 - 1}{X_4} - (1 - \beta)\frac{X_3 - 1}{X_3}}.$$

After some basic algebras we obtain;

$$1 - \beta = \frac{-pX_3X_4^2 - X_1X_2X_3 + pX_1X_3^2 + pX_3^2X_4}{(1 - p)X_3X_4^2 - X_1X_2X_3 + pX_1X_3^2 + X_1X_4^2 - pX_1X_3X_4},$$

then:

$$T_{\beta^*}(F) = \frac{(1 - p)X_3X_4^2 - X_1X_3X_4 + pX_1X_3^2 + X_1X_4^2 - pX_1X_3X_4}{X_3X_4(1 + p) - p(X_3^2 + X_4^2)}.$$

For the WESS $\widehat{\gamma}$, the mean occupation measure in state F is given by:

$$T_{\widehat{\gamma}}(F) = \widehat{\gamma} f_{u_2}(2, h) + (1 - \gamma) f_{u_1}(2, l) = \gamma X_4 + (1 - \gamma)X_3.$$

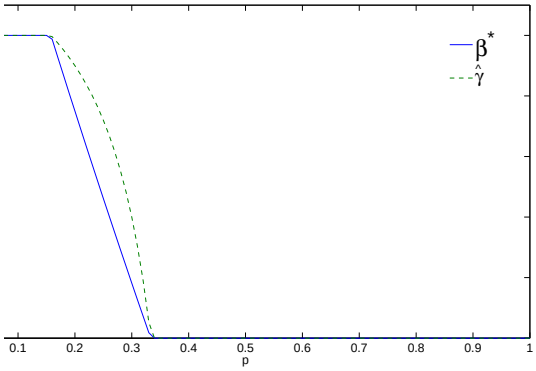
$$1 - \gamma = \frac{pX_4^2 - pX_3X_4 - pX_1X_3 + X_1X_4}{pX_3^2 + pX_4^2 - X_3X_4 - pX_3X_4},$$

which leads:

$$\begin{aligned} T_{\widehat{\gamma}}(F) &= \frac{X_1X_3X_4(1 + p) + (p - 1)X_3X_4^2 - pX_1X_3^2 - X_1X_4^2}{p(X_3^2 + X_4^2) - X_3X_4(1 + p)}, \\ &= T_{\beta^*}(F) \end{aligned}$$

$\diamond$

We compare numerically the parameter β that defines the stationary WESS β^* (i.e. the probability of choosing high power at state 2) with the parameter γ



that defines the mixed WESS $\widehat{\gamma}$ (i.e. the probability to use always high power when at state 2) depending on the probability p in order to validate our results.

We observe on figure 7 that the parameters are equal when they are pure but different otherwise in the mixed case. For example when $p = 0.1$, the probability that a mobile has to be alone is very low, both WESS are equal to one, the mobile uses high power level. When $p = 0.25$, we have $\beta = 0.46$ and $\gamma = 0.72$. Finally, when the probability p is high, in a given scale, both WESS are equal to zero, mobile uses low power level. We remark that the probability γ is larger that the corresponding β whenever they do not correspond to pure policies.

8 Conclusions

In this paper we have studied a new class of evolutionary games which we call MDEG, where the decision of each player determine transition probabilities between individual state. We have illustrated this class of game through an energy control problem in wireless networks. We had introduced already in in [4] a definition of ESS strategies in stationary policies in a particular simple MDEG in which only in one state there are decisions to be taken. If we apply directly that definition to general policies (we call this here a Strong ESS) it turns out that when abandoning the restriction to stationary policies, even in this simple model there are no ESS (except for some restricted choice of parameters that results in some pure ESS). We solved this problem by defining a weaker notion of ESS using occupation measures. We have then proposed methods to determine weak and strong ESS.

References

[1] J.W. Weibull, *Evolutionary Game Theory*, Cambridge, MA: MIT Press, 1995.

[2] Josef Hofbauer and Karl Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press 1998.

[3] L. Shapley, *Stochastic Games*, Proceedings of the National Academy of Sciences of the USA, 39, 1953.

[4] E. Altman, Y. Hayel, *A Stochastic Evolutionary Game Approach to Energy Management in a Distributed Aloha Network*, proceedings of IEEE INFOCOM, 2008.

[5] A. I. Houston and J. M. McNamara, "Fighting for food: a dynamic version of the Hawk-Dove game", *Evolutionary Ecology*, Vol 2, pp. 51-64, 1988.

[6] A. I. Houston and J. M. McNamara, "Evolutionary stable strategies in the repeated hawk-dove game", *Behavioral Ecology*, Vol 2 No 3, pp. 219-227, 1991.

[7] J. McNamara, S. Merad, E. Collins, *The Hawk-Dove Game as an Average-Cost Problem*, Advances in Applied Probability, vol. 23, no. 4, 1991.

[8] E. Altman, *Constrained Markov Decision Process*, Chapman and Hall/CRC, 1999.

[9] Valentina Corradi and Rajiv Sarin, "Continuous Approximiations of Stochastic Evolutionary Game Dynamics", Journal of Economic Theory Volume 94, Issue 2, October 2000, Pages 163-191

[10] Dean Foster and Peyton Yong, "Stochastic Evolutionary game dynamics", *Theoreticial Population Biology*, Vol 38 No 2, October 1990.

[11] Lorens A Imhof, "Long-run behavior of the stochastic replicator dynamics", *Annals of Appl Probability*, Vol 15 No 1B, 1019-1045, 2005.

[12] L. C. M. Kallenberg (1983), *Linear Programming and Finite Markovian Control Problems*, Mathematical Centre Tracts 148, Amsterdam.

[13] Michel Benaïm, Joseph Hofbauer and William H. Sandholm, "Robust Permanence and Impermanence for the Stochastic Replicator Dynamic", available in
http://members.unine.ch/michel.benaim/perso/bhsa.pdf

[14] D. J. White, *Markov Decision Processes*, J. Wiley, 1993.

A Computation of the matrix of the related EG

In this appendix we show the computation of each elements of the standard EG derived from the MDEG of energy control.

For the initial state 2, we have the following probabilities:

$$\forall t>0, \quad p_t(\eta,u_1;2,l)=Q(2,l)^t, \quad \text{and} \quad p_t(\eta,u_1;2,h)=0.$$

$$\forall t>0, \quad p_t(\eta,u_2;2,l)=0, \quad \text{and} \quad p_t(\eta,u_2;2,h)=Q(2,h)^t.$$

$$\forall t>0, \quad p_t(\eta,u_3;2,l)=Q(2,l)^t, \quad \text{and} \quad p_t(\eta,u_3;2,h)=0.$$

$$\forall t>0, \quad p_t(\eta,u_4;2,l)=0, \quad \text{and} \quad p_t(\eta,u_4;2,h)=Q(2,h)^t.$$

For the intermediate state 1, we have the following probabilities:

$$\forall t>0, \quad p_t(\eta,u_1;1,l)=(1-Q(2,l))\sum_{j=0}^{t-1}Q(1,l)^{t-1-j}Q(2,l)^j, \quad \text{and} \quad p_t(\eta,u_1;1,h)=0.$$

$$\forall t>0, \quad p_t(\eta,u_2;1,l)=(1-Q(2,h))\sum_{j=0}^{t-1}Q(1,l)^{t-1-j}Q(2,h)^j, \quad \text{and} \quad p_t(\eta,u_2;1,h)=0.$$

$$\forall t>0, \quad p_t(\eta,u_3;1,l)=0, \quad \text{and} \quad p_t(\eta,u_3;1,h)=(1-Q(2,l))\sum_{j=0}^{t-1}Q(1,h)^{t-1-j}Q(2,l)^j.$$

$$\forall t>0, \quad p_t(\eta,u_4;1,l)=0, \quad \text{and} \quad p_t(\eta,u_4;1,h)=(1-Q(2,h))\sum_{j=0}^{t-1}Q(1,h)^{t-1-j}Q(2,h)^j.$$

For the absorbent state 0 the probability to take an action is null because no actions are available is this state.

Now, we are able to compute the occupation measure through the expected number of time units $f_{\eta,u}(s,a)$ during which a user is in state s and it chooses action $a$corresponding to a policy u. For the policy $u_1=(l,l)$, we have:

$$f_{\eta,u_1}(2,l)=\frac{1}{1-Q(2,l)}, \quad f_{\eta,u_1}(2,h)=0, \quad f_{\eta,u_1}(1,h)=0,$$

and

$$\begin{aligned} f_{\eta,u_1}(1,l) &= (1-Q(2,l))\sum_{t=1}^{+\infty}\sum_{j=0}^{t-1}Q(1,l)^{t-1-j}Q(2,l)^j, \\ &= (1-Q(2,l))\left[\sum_{t=0}^{+\infty}Q(1,l)^t\sum_{t=0}^{+\infty}Q(2,l)^t\right], \\ &= \frac{1}{1-Q(1,l)}. \end{aligned}$$

In the same way, we have for the policy $u_2=(l,h)$:

$$f_{\eta,u_2}(2,l)=0, \quad f_{\eta,u_2}(2,h)=\frac{1}{1-Q(2,h)}, \quad f_{\eta,u_2}(1,h)=0,$$

and

$$\begin{aligned} f_{\eta,u_2}(1,l) &= (1-Q(2,h))\sum_{t=1}^{+\infty}\sum_{j=0}^{t-1} Q(1,l)^{t-1-j}Q(2,h)^j, \\ &= (1-Q(2,h))\left[\sum_{t=0}^{+\infty} Q(1,l)^t \sum_{t=0}^{+\infty} Q(2,h)^t\right], \\ &= \frac{1}{1-Q(1,l)}. \end{aligned}$$

We have similarly for the policy $u_3 = (h,l)$:

$$f_{\eta,u_3}(2,l) = \frac{1}{1-Q(2,l)}, \quad f_{\eta,u_3}(2,h) = 0, \quad f_{\eta,u_3}(1,l) = 0,$$

and

$$\begin{aligned} f_{\eta,u_3}(1,h) &= (1-Q(2,l))\sum_{t=1}^{+\infty}\sum_{j=0}^{t-1} Q(1,h)^{t-1-j}Q(2,l)^j, \\ &= (1-Q(2,l))\left[\sum_{t=0}^{+\infty} Q(1,h)^t \sum_{t=0}^{+\infty} Q(2,l)^t\right], \\ &= \frac{1}{1-Q(1,h)}. \end{aligned}$$

Finally, for the policy $u_4 = (h,h)$:

$$f_{\eta,u_4}(2,l) = 0, \quad f_{\eta,u_4}(2,h) = \frac{1}{1-Q(2,h)}, \quad f_{\eta,u_4}(1,l) = 0,$$

and

$$\begin{aligned} f_{\eta,u_4}(1,h) &= (1-Q(2,h))\sum_{t=1}^{+\infty}\sum_{j=0}^{t-1} Q(1,h)^{t-1-j}Q(2,h)^j, \\ &= (1-Q(2,h))\left[\sum_{t=0}^{+\infty} Q(1,h)^t \sum_{t=0}^{+\infty} Q(2,h)^t\right], \\ &= \frac{1}{1-Q(1,h)}. \end{aligned}$$

The modified fitness function $\widetilde{F}$ defined in equation 8 is expressed by:

$$\begin{aligned}
\widetilde{F}(v,u) &= \sum_{s,a} f_{\eta,v}(s,a) \sum_{s',a'} f_{\eta,u}(s',a') r(s,a,s',a'), \\
&= p \sum_{s,a} f_{\eta,v}(s,a) \sum_{s',a'} f_{\eta,u}(s',a') + (1-p) \sum_{s} f_{\eta,v}(s,h) \sum_{s'} f_{\eta,u}(s',l), \\
&= p \sum_{s,a} f_{\eta,v}(s,a) \sum_{s',a'} f_{\eta,u}(s',a') + (1-p) H(v).L(u), \\
&= p(H(v)+L(v))(H(u)+L(u)) + (1-p) H(v).L(u),
\end{aligned}$$

with

$$H(v) = f_{\eta,v}(1,h) + f_{\eta,v}(2,h) \quad \text{and} \quad L(u) = f_{\eta,u}(1,l) + f_{\eta,u}(2,l).$$

Note that $H(u_1) = L(u_4) = 0$.