General Procedures Leading to Correlated Equilibria[*]

Amotz Cahn[†]
Center for the Study of Rationality and Department of Mathematics
The Hebrew University of Jerusalem
91904 Jerusalem, Israel

May 11, 2004

Abstract

Hart and Mas-Colell [2000] show that if all players play "regret-matching" strategies, i.e., they play with probabilities proportional to the regrets, then the empirical distribution of play converges to the set of correlated equilibria, and the regrets of every player converge to zero. Here we show that if only one player, say player i, plays with these probabilities, while the other players are "not too sophisticated," then the result that player i's regrets converge to zero continues to hold. The condition of "not too sophisticated" essentially says that the effect of one change of action of player i on the future actions of the other players decreases to zero as the horizon goes to infinity. Furthermore, we generalize all these results to a whole class of "regret-based" strategies introduced in Hart and Mas-Colell [2001]. In particular, these simplify the "conditional smooth fictitious play" of Fudenberg and Levine [1999].

1. Introduction

A game G of N players is a triplet $\left\langle N, \left\{S^i\right\}_{i\in N}, \left\{u^i\right\}_{i\in N}\right\rangle$, where N is the set of players, S^i is the set of strategies of player i, and $u^i : \prod_{j\in N} S^j \to \mathbb{R}$ is the payoff function of player i. All sets of players and strategies are finite. Denote by $S := \prod_{i\in N} S^i$ the set of strategies of all players, and by $S^{-i} := \prod_{j\neq i, j\in N} S^j$ the set of strategies of all players different from i. Denote by ΔS^i the set of probabilities on S^i (similarly ΔS is the set of probabilities on S). A strategy $s^i \in S^i$ is a pure strategy of player i, and the elements of ΔS^i are the mixed

[*] This is a revision of the author's M.Sc. thesis, May 2000.

[†] The author thanks Professor Sergiu Hart for his help and guidance, and the Associate Editor and an anonymous referee for their comments. I am grateful to my parents and wife for everything.

strategies of player i. When dealing with mixed strategies one often is interested not in the actual payoff received, but rather in the expected payoff using those mixed strategies.

One can also consider a situation where the game G is repeated over and over again. In this situation the strategy the players play at time t is denoted s_t (and the strategy of player i is correspondingly s_t^i). In this paper we study repeated games, and their relations to the solution concepts of the one-shot game.

1.1. Correlated Equilibria

In a two-player zero-sum game there exists the value of the game, which is the payoff that the players can ensure they will get, and optimal strategies, which ensure this payoff. In a non-zero-sum game and in a game with more than two players, we cannot talk about a value of the game since such a value does not exist. Instead of value and optimal strategies one considers equilibrium. Equilibrium is a vector of strategies, such that no player will increase his payoff by unilaterally changing his strategy. The leading non-cooperative equilibrium notion for N-person games in strategic (normal) form is Nash equilibrium, which is an N-tuple of probabilities on S^i, such that no player can increase his payoff by unilaterally changing his action.

The notion of Nash equilibrium has been generalized by Aumann [1974], who introduced the concept of *correlated equilibrium.* Assume that, before the game is played, every player receives a private signal (which does not affect the payoffs). The player may (but need not) choose his action in the game depending on this signal. A correlated equilibrium of the original game is just a Nash equilibrium of the game with the signals. If the signals are (stochastically) independent across the players, this is just a Nash equilibrium (in mixed or pure strategies) of the original game. But the signals could well be correlated, in which case new equilibria may obtain.

Equivalently, a correlated equilibrium is a probability distribution on N-tuples of actions, which can be interpreted as the distribution-of-play instructions given to the players by some “device” or “referee.” Every player is given — privately — instructions for his play only; the joint distribution is known to all of them. Also, for every possible instruction that a player receives, the player realizes that the instruction provides a best response to the random estimated play of the other players — assuming they all follow their instructions.

Finally, one can think of the set of correlated equilibria as the following subset of ΔS, the set of probability distributions on N-tuples of actions: $x \in \Delta S$ is a correlated equilibrium if for any random variable $Y = \left(Y^i\right)_{i \in N}$ (Y^i with values in S^i) such that[1] $Y \sim x$, the following holds for all $i \in N$, and for all $s^i \in S^i$ such that $\Pr\left(Y^i = s^i\right) > 0$:

$$\mathrm{E}\left[u^i\left(s^i, Y^{-i}\right) \mid Y^i = s^i\right] = \max_{j \in S^i} \mathrm{E}\left[u^i\left(j, Y^{-i}\right) \mid Y^i = s^i\right]. \tag{1.1}$$

1.2. Regrets

A player may be sorry because he played one way instead of another. We can quantify how sorry he is by the difference between the payoff he would have gotten had he played differently, and the payoff he actually received. This difference between payoffs is called the regret.

The regret we consider is obtained by comparing the actual payoff received to the payoff one would have gotten by playing another pure strategy. In a one-shot game this regret does not make much sense when dealing with mixed strategies. However, when dealing with a repeated game the picture is different. In a repeated game this regret can be described for player i as the difference between the average payoff of playing k instead of j every time player i played strategy j, and the average payoff of the actually played strategies. We denote this regret by $D_t^i(j,k)$ — the *regret at time t of player i from j to*[2] *k.* This regret is of course a function of how often every strategy $s \in S$ was played, or, put differently, a function of the proportion of play of every strategy $s \in S$. This proportion is called the *empirical distribution of play,* and it depends on the time t of the game. It is denoted by[3] z_t. Notice that z_t is a probability distribution

[1] The notation $Y \sim x$ means that Y and x have the same probability distribution.

[2] A formula for this is

$$D_t^i(j,k) = \frac{1}{t} \sum_{\tau \le t; s_\tau^i = j} \left[u^i\left(k, s_\tau^{-i}\right) - u^i\left(j, s_\tau^{-i}\right)\right].$$

[3] $z_t(s)$ equals the number of times s was actually played in the first t periods, divided by t.

on S, the set of all strategies.[4]

Hart and Mas-Colell [2000] (henceforth [**HM1**]) found an interesting connection (see Section 3 there) between regrets and the set of correlated equilibria (which can be described as a subset of the set of probability distributions on S). They prove: Given any $\varepsilon \geq 0$, let $\{s_t\}_{t=1,2,\ldots}$ be a sequence of plays such that the *limsup* of the regret for every player and every strategy is less than or equal to ε. Then the sequence of empirical distributions z_t converges to the set of correlated ε-equilibria. Furthermore, Hart and Mas-Colell [HM1, Theorem A] use Blackwell's [1956] Approachability Theorem to prove that the set of all nonpositive regret vectors is an approachable set for every player. This means that every player has an adaptive strategy such that, no matter what the other players do, all his regrets converge to the nonpositive orthant. However, this strategy is quite complicated, and one must calculate eigenvectors of a different matrix in every period time t in order to evaluate it. Therefore Hart and Mas-Colell [HM1, Section 2] construct a simple adaptive procedure in which the transition probabilities are linearly proportional to the regrets. This procedure has the property that if all players follow it, then the regrets of every player converge to the nonpositive orthant. (Hence the empirical distribution z_t converges to the set of correlated equilibria.) Nevertheless, as they state, this property holds only if all players follow this procedure. Here (in Section 3) we give weak conditions on the other players' play that suffice for the regrets of player i to converge to the nonpositive orthant.[5] Furthermore, following Hart and Mas-Colell [2001] (henceforth [**HM2**]), we generalize (in Section 4) the Hart–Mas-Colell strategy to a larger class of strategies and we prove that the same convergence theorems hold. In particular, in Section 4.5, we strengthen a theorem of Fudenberg and Levine dealing with conditional smooth fictitious play.

[4] A way of describing the regret of player i from j to k in terms of z_t is

$$D_t^i(j,k) = \sum_{s \in S : s^i = j} z_t(s) \left[u^i\left(k, s^{-i}\right) - u^i(s) \right].$$

[5] Note that a procedure that is totally correlated between players is not of major interest. Players can decide on any Nash equilibrium, and this would be a strategy that leads to equilibria. Hart and Mas-Colell's procedure is not such a procedure, since it is not based on the game played. Yet, our result gives added importance to this procedure.

2. Preliminaries

2.1. The Hart–Mas-Colell Simple Procedure

As mentioned in the introduction Hart and Mas-Colell develop a simple method that ensures that the empirical distribution of play will converge with probability one to the set of correlated equilibria.

We describe here the procedure that they developed.

Let G be a game with a finite number of players. Suppose that G is played repeatedly through time: $t = 1, 2, 3, \ldots$. Let s_t^i be the strategy that player i played at time t (and s_t^{-i} the strategy other players played, and s_t the strategy combination of all players at time t). Let $h_t := (s_\tau)_{\tau \leq t}$ be the history of the game until time t.

Let

$$A_t^i(j,k) := 1_{\left\{s_t^i = j\right\}} \left[u^i\left(k, s_t^{-i}\right) - u^i(s_t)\right]$$

be the regret at the specific time t from j to k. The regret $D_t^i(j,k)$ is the average of $A_\cdot^i(j,k)$, i.e.,

$$D_t^i(j,k) = \frac{1}{t} \sum_{\tau=1}^{t} A_\tau^i(j,k).$$

The positive part of the regret, denoted $R_t^i(j,k)$, is $R_t^i(j,k) := \left[D_t^i(j,k)\right]_+$. As we mentioned the transition probabilities $\pi_t^i(j,k)$ of the *Hart–Mas-Colell strategy* (henceforth **HMS**) are proportional to the positive part of the regret. Let μ be sufficiently large,[6] let $\pi_t^i(j,k) := (1/\mu) \cdot R_t^i(j,k)$ for $k \neq j \in S^i$, and $\pi_t^i(j,j) := 1 - \sum_{k \in S^i : k \neq j} \pi_t^i(j,k)$.

In the Hart and Mas-Colell strategy player i plays at time $t+1$ according to the probabilities

$$\Pr\left(s_{t+1}^i = s^i \mid h_t\right) = \pi_t^i\left(s_t^i, s^i\right); \tag{2.1}$$

that is, the transition probabilities from one period to the next are linearly proportional to the regrets. Hart and Mas-Colell prove [HM1, Section 2] that if all players follow this strategy then with probability one the empirical distribution of play converges to the set of correlated equilibria. The method used in order to prove this is to show that the positive part of the regret converges to zero almost

[6] Specifically, μ should be large enough to ensure $\pi_t^i(j,j) > 0$ for every $i \in N$ and every $j \in S^i$.

surely for every player and every strategy. (Henceforth, whenever we use the term "regret converges to zero," we mean the positive part, i.e., $R^i_t\left(j,k\right) \to_{t\to\infty} 0$ for all $k \neq j \in S^i$.)

However, if other players do not follow HMS then the regret of player i need not converge to zero. In Section 3 we show that with some slight conditions on the other players' play we can still get this convergence.

2.2. Approachable Sets

Consider a game in strategic form played by a player i against an opponent $-i$ (which can be nature, or another player, or many other players). The action sets are the finite sets S^i for i and S^{-i} for $-i$. The payoff functions are vectors in some Euclidean space. Let a_t be the payoff to player i at time t, and $\overline{a}_t$ be the average payoff to player i up to time t. A set C is called an approachable set for player i if player i can guarantee, no matter what player $-i$ does, that the Euclidean distance, $\operatorname{dist}\left(\overline{a}_t, C\right)$, tends to zero almost surely as $t \to \infty$.

Given a game with a scalar payoff u^i, we can look at the vector of regrets of player i in the one-shot game A^i (defined in the previous subsection) as a vector payoff. Now we can consider this vector payoff and ask which sets are approachable. Hart and Mas-Colell [HM1, Section 3] prove that the nonpositive orthant (denoted $\mathbb{R}^{m_i}_-$ where $m_i = \left|S^i\right|$) is approachable for every player i. (Obviously, other sets are also approachable, e.g., any set that includes $\mathbb{R}^m_-$, which may correspond to correlated ε-equilibria.)

Consider a convex closed set C such that $\mathbb{R}^m_- \subseteq C$ and a mapping $\Lambda : \mathbb{R}^m \backslash C \to \mathbb{R}^m$ such that Λ is continuous, integrable, and, for every $x \in \mathbb{R}^m \backslash C$, the vector $\Lambda\left(x\right)$ represents a direction from C to x, in the following sense: $\Lambda\left(x\right) \cdot x > \Lambda\left(x\right) \cdot y$ for all $y \in C$. Hart and Mas-Colell [HM2] prove that if a player uses a strategy which guarantees that the one-shot payoffs lie in the C-side of the half space generated by $\Lambda\left(x\right)$ (and not in the x-side), then the average payoff will converge to C almost surely [HM2, Theorem 2.1]. In Section 4 we prove, similarly to what Hart and Mas-Colell do in their simple adaptive procedure, that we can use a strategy linearly proportional to the direction pointed out by operating Λ on the regret (that is, $\Lambda\left(D^i_t\right)$), to get, if all players follow this strategy, a similar result. Furthermore, we show that with some slight conditions on other players' play, we can get the convergence of the regret for player i to this set C, no matter what the other players do.

3. Consistency of the Hart–Mas-Colell Strategy

3.1. Introduction

In their paper [HM1], Hart and Mas-Colell show a simple adaptive procedure leading to correlated equilibrium, as we described in Section 2.1. However, as they point out in Section 4(d), this procedure is not "conditionally universally consistent."[7] In particular, if only player i follows the procedure we cannot conclude that all his regrets go to zero. In this section we give sufficient conditions on the behavior of the other players, which imply that all regrets of player i will necessarily converge to zero.

3.2. Main Results of This Section

Let G a be a game with a finite number of players. When dealing with player i we can always look at G as a two-player game between i and $-i$, where $-i$ is $N\backslash\{i\}$.[8] For any strategy used by $-i$, we can ask: what is the effect of the action actually used by player i at stage t of the game, on the action player $-i$ uses at step $t+w$? We show that if this effect is as small as $f(w)/g(t)$, for some functions f,g such that $g(t)\rightarrow_{t\rightarrow\infty}\infty$, and player i uses the Hart and Mas-Colell strategy HMS (2.1), then, no matter what strategy player $-i$ uses, the regrets of player i will converge to zero as time goes to infinity.

Formally, assume that for all $t,w>0$, given two histories h_{t+w-1} and h'_{t+w-1} such that for every $\tau<t+w$, $\tau\neq t$ we have $s_\tau=s'_\tau$ and for $\tau=t$ we have $s_\tau^{-i}=s_\tau'^{-i}$, $s_\tau^i\neq s_\tau'^i$ (that is, the two histories h_{t+w-1} and h'_{t+w-1} differ only in player i's action at time t), then for all s^{-i} in S^{-i}

$$\left|\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t+w-1}\right)-\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h'_{t+w-1}\right)\right|\leq\frac{f(w)}{g(t)} \tag{3.1}$$

for some functions f,g such that $g(t)\rightarrow_{t\rightarrow\infty}\infty$. (Note that there are no conditions on f; its role is to get a uniform bound for every w.)

[7] The result is not guaranteed for a player unless all players play according to this strategy.

[8] We allow other players to be correlated among themselves; hence we may refer to them as one player.

What this condition says is that the effect of one change in the action of i on the action of the other players converges to zero as the horizon goes to infinity. (For example in the HMS this is so since the effect is of the order $1/t$.)

Remarks:

(1) In fact, we need this condition only for w such that $w = o(t)$.

(2) The interdependence between the strategies of the players other than i is irrelevant.

(3) If all players follow HMS as in (2.1), then by Step $M3$ of the Appendix of [HM1] $R^i_{t+w}(j,k) - R^i_t(j,k) = O(w/t)$ (and the same holds for the corresponding transition probabilities); hence (3.1) holds with f a linear function of w (specifically, $f(w) = 2cw$ where c is the constant for $O(w/t)$ and $g(t) = t$).

Theorem 3.1. *If player i uses the Hart–Mas-Colell simple strategy of (2.1), then the regrets of player i are guaranteed to converge to zero a.s. as $t \to \infty$, for any strategies of the other players that satisfy (3.1).*

3.3. Proof of Theorem

We shall make use of the following remarks.

3.3.1. Remarks

In (3.1) we may assume without loss of generality that:

(*1) $f(w) \geq w$; and f is monotone increasing. (Otherwise, take $F(w) := w + \max_{w' \leq w} f(w')$ instead of f.) We extend f to the entire positive real line in a manner which will make it one-to-one.

(*2) $g(t) \leq t$, $(g(t) \geq 1)$ and g is monotone nondecreasing. (Otherwise, define $G(t) := \min\{t, \min_{x \geq t} g(x)\}$, and take G instead of g.)

(*3) $g(t) - g(t-1) \leq 1/t$ (Otherwise, one can define $\hat{G}(t)$ by $\hat{G}(1) := g(1)$ and $\hat{G}(t+1) := \min\left\{g(t+1), \hat{G}(t) + 1/(t+1)\right\}$, satisfying $\hat{G}(t) \to_{t\to\infty} \infty$ and $\hat{G}(t) - \hat{G}(t-1) \leq 1/t$, and take $\hat{G}(t)$ instead of g. [$\hat{G}$ satisfies (*1) and (*2) if g does.])

Notice that we require that (3.1) hold for every history. Thus, assume that we have two histories h_{t+w-1} and h'_{t+w-1} such that for every $\tau < t$ we have $s_\tau = s'_\tau$ and for $\tau \geq t$ we possibly have $s^i_\tau \neq s'^i_\tau$, $s^{-i}_\tau = s'^{-i}_\tau$. Then we can define a sequence of histories h_{t+w-1}^{0} , h_{t+w-1}^{1} ,..., h_{t+w-1}^{w} , such that $h_{t+w-1}^{0}=h_{t+w-1}$, and h_{t+w-1}^{l} differs from h_{t+w-1}^{l+1} only in that at stage $t+l$, we have $h_{t+w-1}^{l+1}\left(s_{t+l}\right)=s'_{t+l}$. Hence $h_{t+w-1}^{w}=h'_{t+w-1}$, and

$$\begin{aligned}\left|\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t+w-1}\right)-\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h'_{t+w-1}\right)\right| &\leq \\ \sum_{l=0}^{w-1}\left|\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t+w-1}^{l}\right)-\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t+w-1}^{l+1}\right)\right| &\leq \sum_{l=0}^{w-1}\frac{f\left(w-l\right)}{g\left(t+l\right)}.\end{aligned}$$

Now since we have assumed that f is monotone increasing, and g is monotone nondecreasing, it follows that:

$$\left|\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t+w-1}\right)-\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h'_{t+w-1}\right)\right|\leq\frac{wf\left(w\right)}{g\left(t\right)}.$$

If we define $f^{*}\left(w\right):=wf\left(w\right)$ instead of $f\left(w\right)$ we get the following: there exists f,g such that $g\rightarrow\infty$, and for every h_{t+w-1} and h'_{t+w-1} such that for every $\tau<t$ we have $s_{\tau}=s'_{\tau}$ and for $\tau\geq t$ we possibly have $s_{\tau}^{i}\neq s_{\tau}^{\prime i}$, $s_{\tau}^{-i}=s_{\tau}^{\prime -i}$

$$\left|\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t+w-1}\right)-\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h'_{t+w-1}\right)\right|\leq\frac{f\left(w\right)}{g\left(t\right)} \tag{3.2}$$

holds. Henceforth, we assume that the strategy of $-i$ satisfies (3.2), with f and g satisfying (*1)-(*3).

An analogous way of looking at this situation is as follows. If all players follow HMS, then there exists a matrix of transition probabilities $\Pi_{t}^{i'}$, for every player i' and stage t. For every player i' we have $\left|\Pi_{t}^{i'}-\Pi_{t+w}^{i'}\right|=O\left(w/t\right)$; hence

$$\left|\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t+w-1}\right)-\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t},s_{t+1}^{-i},...,s_{t+w-1}^{-i}\right)\right|=O\left(\frac{w}{t}\right).$$

Consider a given process s such that player i follows HMS and the other players do not. Suppose that

$$\left|\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t+w-1}\right)-\Pr\left(s_{t+w}^{-i}=s^{-i}\mid h_{t},s_{t+1}^{-i},...,s_{t+w-1}^{-i}\right)\right|=O\left(\frac{f\left(w\right)}{g\left(t\right)}\right)$$

for some functions f,g such that $g\left(t\right)\rightarrow_{t\rightarrow\infty}\infty$. We prove that the regrets of i converge to zero almost surely.

3.3.2. Lemma

Before proving the theorem we state a simple lemma.

Lemma 3.2. *Assume that the players $-i$ are using strategies that are independent of player i's moves, that is,*

$$\Pr\left(s_t^{-i} = s^{-i} \mid h_{t-1}\right) = \Pr\left(s_t^{-i} = s^{-i} \mid s_1^{-i}, ..., s_{t-1}^{-i}\right),$$

and that player i uses HMS as in (2.1). Then the regrets of player i converge to zero a.s., as $t \to \infty$.

Proof: The proof is exactly as in [HM1], except that in Step $M4$ here we define $\hat{s}_{t+w}$ differently. $\hat{s}_{t+w}$ is defined by $\hat{s}_t := s_t$, and the transition probabilities are:

$$\Pr\left(\hat{s}_{t+w} = s \mid \hat{s}_t, ..., \hat{s}_{t+w-1}\right) = \Pi_t^i\left(\hat{s}_{t+w-1}^i, s^i\right) \cdot \Pr\left(s_{t+w}^{-i} = s^{-i} \mid h_t, \hat{s}_{t+1}^{-i}, ..., \hat{s}_{t+w-1}^{-i}\right).$$

One can verify that Step $M4$ is still true with the same proof given in [HM1]. (Given h_t, player $-i$ plays with the same probabilities for $\hat{s}$ and s; therefore

$$\left|\Pr\left(\hat{s}_{t+w}^{-i} = s^{-i} \mid h_{t+w-1}\right) - \Pr\left(s_{t+w}^{-i} = s^{-i} \mid h_{t+w-1}\right)\right| = 0;$$

and for player i we still have

$$\left|\Pr\left(\hat{s}_{t+w}^{i} = s^{i} \mid h_{t+w-1}\right) - \Pr\left(s_{t+w}^{i} = s^{i} \mid h_{t+w-1}\right)\right| = O\left(\frac{w}{t}\right);$$

hence one can use the same proof.) A similar statement is true also for Step $M5$ of [HM1]. Step $M6$ is also true, and the proof remains unchanged. (Notice that in the proof of $M6$ we are using only the independence of i and $-i$, which still holds, and the fact that player i uses Π_t^i as transition probabilities.) The last step involving the $\hat{s}$ strategies is $M7$, which involves only the stochastic matrix Π_t^i, and it is obviously still true; therefore the continuation of the proof as in [HM1] is still valid. ■

We now can prove our theorem.

3.3.3. Proof of Theorem 3.1

This proof follows the Hart and Mas-Colell paper [HM1], and we use the same steps. We use lowercase letters to distinguish steps in our proof from those in theirs.

- *Steps* $M1, M2, M3$ of [HM1] are generally true , independently of the strategies used.

Define $\hat{s}_{t+w}$ as in the proof of Lemma 3.2.

- *Step m4:* $\left|\Pr\left(\hat{s}_{t+w} = s \mid h_t\right) - \Pr\left(s_{t+w} = s \mid h_t\right)\right| = O\left(\frac{wf(w)}{g(t)}\right)$.

We use Hart and Mas-Colell's lemma in the Proof of Step $M4$ [HM1]. Since for every player, the transition probability for the $\hat{s}$ process differs from the corresponding one for the s process by at most $O\left(f\left(w\right)/g\left(t\right)\right)$ (for player i it differs by $O\left(w/t\right)$ which is $\leq O\left(f\left(w\right)/g\left(t\right)\right)$ by (*1) and (*2)), it follows that

$$\left|\Pr\left(\hat{s}_{t+w} = s \mid h_t\right) - \Pr\left(s_{t+w} = s \mid h_t\right)\right| = \sum_{w' \leq w} O\left(\frac{f\left(w'\right)}{g\left(t\right)}\right) = O\left(\frac{wf\left(w\right)}{g\left(t\right)}\right).$$

(The last equality follows since f is increasing.) □

- *Step m5:* $\left|\alpha_{t,w}\left(j, s^{-i}\right) - \widehat{\alpha}_{t,w}\left(j, s^{-i}\right)\right| = O\left(\frac{wf(w)}{g(t)}\right)$.

This is immediate by Step $m4$.

- *Step m6:* $\widehat{\alpha}_{t,w}\left(j, s^{-i}\right) = \Pr\left(\hat{s}_{t+w}^{-i} = s^{-i} \mid h_t\right)\left[\Pi_t^{w+1} - \Pi_t^{w}\right]\left(s_t^i, j\right)$.

The proof of [HM1, Step *M6*] holds, since, by definition of the $\hat{s}$ process, the transitions of i and $-i$ are independent.

- *Step m7:* $\widehat{\alpha}_{t,w}\left(j, s^{-i}\right) = O\left(w^{-1/2}\right)$.

The proof is the same as in [HM1, Step *M7*].

- *Step m8:* $E\left[\left(t+v\right)^2 \rho_{t+v} \mid h_t\right] \leq t^2\rho_t + O\left(\frac{tv^2 f(v)}{g(t)} + tv^{1/2}\right)$.

By steps $m5, m7$, and $M2$, it follows that

$$\sum_{w=1}^{v} R_t \cdot E\left[A_{t+w} \mid h_t\right] = \sum_{w=1}^{v} O\left(\frac{wf\left(w\right)}{g\left(t\right)} + w^{-1/2}\right) = O\left(\frac{v^2 f\left(v\right)}{g\left(t\right)} + v^{1/2}\right).$$

Substituting into $M1(i)$ yields the result. (Note that the term $O\left(v^2\right)$ is not needed since $tv^2 f\left(v\right)/g\left(t\right) \geq v^2$.) □

Now let t_n be an increasing sequence of positive integers (to be defined later), and let $v_n := t_{n+1} - t_n$. Then

- *Step $m9.1$:* $E\left[t_{n+1}^2\rho_{t_{n+1}} \mid h_{t_n}\right] \leq t_n^2\rho_{t_n} + O\left(\frac{t_n v_n^2 f(v_n)}{g(t_n)} + t_n v_n^{1/2}\right)$.

Step $m9.1$ follows immediately from $m8$.

Let $\widetilde{f}(w) := w^2 f(w)$. (Notice that $\widetilde{f}$ is a continuous strictly increasing function, and thus $\widetilde{f}$ has an inverse function, denoted by $\widetilde{f}^{-1}$.)

Let $a_n := \frac{1}{2}\widetilde{f}^{-1}(g(n))$, and let $t_n := \lceil na_n \rceil$; ($v_n = t_{n+1} - t_n$).

- *Step $m9.2$:*

(i) a_n is a nondecreasing sequence, and $a_n \rightarrow_{n\rightarrow\infty} \infty$.
(ii) $\frac{v_n}{t_n} = O\left(n^{-1}\right)$.
(iii) $\widetilde{f}(v_n)/g(t_n) = O(1)$.
(iv) $\frac{t_n v_n^2 f(v_n)}{g(t_n)} + t_n v_n^{1/2} = O\left(n a_n^{3/2}\right)$.

Proof: *(i)* is immediate since all the functions involved are increasing and go to infinity.

(ii) By (*3) we have $g(t) - g(t-1) \leq 1/t$; also, for $x \geq y \geq 1$ we have $\widetilde{f}^{-1}(x) - \widetilde{f}^{-1}(y) \leq x - y$ since f is increasing and greater than or equal to 1 by (*1). Hence

$$2(a_n - a_{n-1}) \leq g(n) - g(n-1) \leq \frac{1}{n}.$$

Thus

$$\begin{aligned}\frac{v_n}{t_n} &\leq \frac{1 + (n+1)a_{n+1} - na_n}{na_n} \\ &\leq \frac{1 + (n+1)\left(a_n + \frac{1}{2(n+1)}\right) - na_n}{na_n} = \frac{a_n + \frac{3}{2}}{na_n} = O\left(\frac{1}{n}\right),\end{aligned}$$

as claimed. □

(iii) By *(i)* there exists an n_0 such that $a_n \geq 3/2$ for all $n > n_0$. First, we have $v_n \leq 2a_n$ for all $n > n_0$. Indeed,

$$\begin{aligned}v_n - 2a_n &\leq 1 + (n+1)a_{n+1} - na_n - 2a_n \\ &\leq 1 + (n+1)\left(a_n + \frac{1}{2(n+1)}\right) - (n+2)a_n = 1.5 - a_n \leq 0.\end{aligned}$$

Thus $\widetilde{f}(v_n) \leq \widetilde{f}(2a_n) = g(n) \leq g(t_n)$ since $t_n \geq n \cdot 1.5$, which yields the result.□

(iv) By *(iii)*, the first term is $O(t_n)$; thus, in total we have $O\left(t_n v_n^{1/2}\right)$, which by *(ii)* is $O\left(t_n (t_n/n)^{1/2}\right) = O\left(n a_n^{3/2}\right)$. □

- *Step* $m10$*:* $\lim_{n\to\infty}\rho_{t_n}=0$ a.s.

Proof: Define $b_n:=t_n^2\approx n^2a_n^2$ and $X_n:=b_n\rho_{t_n}-b_{n-1}\rho_{t_{n-1}}=t_n^2\rho_{t_n}-t_{n-1}^2\rho_{t_{n-1}}$.

By Step $M1(ii)$ it follows that $|X_n|\le O\left(t_nv_n+v_n^2\right)$, which by $M9.2(ii)$ equals $O\left(t_n^2/n\right)$. Thus

$$\sum_n\frac{1}{b_n^2}\operatorname{Var}\left(X_n\right)=\sum_n O\left(\frac{1}{n^2}\right)<\infty.$$

Next, Steps $m9.1$ and $m9.2$ imply that

$$\begin{aligned}\frac{1}{b_n}\sum_{\nu\le n}E\left[X_\nu\mid X_1,...,X_{\nu-1}\right] &= O\left(n^{-2}a_n^{-2}\right)\cdot\sum_{\nu\le n}O\left(\frac{t_\nu v_\nu^2 f\left(v_\nu\right)}{g\left(t_\nu\right)}+t_\nu v_\nu^{1/2}\right)\\ &= O\left(n^{-2}a_n^{-2}\right)\cdot\sum_{\nu\le n}O\left(\nu a_\nu^{1.5}\right)=O\left(a_n^{-0.5}\right)\to_{n\to\infty}0.\end{aligned}$$

(For the first equality we use $m9.1$, for the second equality we use $m9.2(iv)$, and for the third we use $a_\nu\le a_n$ for $\nu\le n$ which follows by $m9.2(i)$, and $\sum_{\nu\le n}\nu=O\left(n^2\right)$.)

Applying the Strong Law of Large Numbers for Dependent Random Variables yields that ρ_{t_n}, which is nonnegative and equals $\left(1/b_n\right)\sum_{\nu\le n}X_\nu$, must converge to 0 a.s. (Here and in the sequel, whenever we mention the Strong Law of Large Numbers for Dependent Random Variables, we refer to Theorem 32.1.E of Loève [1978], also quoted in Step $M10$ of [HM1].) □

- *Step* $m11$:. $\lim_{t\to\infty}R_t\left(j,k\right)=0$ a.s.

The proof in [HM1] applies. ■

3.4. Counterexample

We exhibit an example where the conditions fail and the regrets do not converge to zero.

Let $M>0$ be as large as desired, and $1>\rho>0$ be as small as desired. We show that even if we demand that player i's action have no effect on the strategy of player $-i$ for the following $M-1$ periods after it was played, and that after M periods the effect be no larger than ρ, yet we can construct an example where the regrets of player i will not converge to zero almost surely.

Example 3.3. *Consider a two-player game in which the payoff matrix of player* 1 *is:*

$$\begin{array}{c|c|c|}
 & L & R \\ \hline
T & \mathit{1} & \mathit{0} \\ \hline
B & \mathit{0} & \mathit{1} \\ \hline
\end{array}$$

Let $M > 0$ be as large as desired, and $1 > \rho > 0$ be as small as desired. Let player 1 use HMS (2.1), and let player 2 use the following strategy:

$$\Pr\left[s_t^2 = R \mid h_t\right] = \begin{cases} 1 & \text{if } s_{t-1}^2 = R \text{ and } s_{t-M}^1 = T \\ 1-\rho & \text{if } s_{t-1}^2 = R \text{ and } s_{t-M}^1 = B \\ \rho & \text{if } s_{t-1}^2 = L \text{ and } s_{t-M}^1 = T \\ 0 & \text{if } s_{t-1}^2 = L \text{ and } s_{t-M}^1 = B \end{cases}$$

for all $t > M$ (and arbitrary for $t \leq M$). That is, player 2 changes his action with probability ρ to the worst reply (from the point of view of player 1) to player 1's action M periods ago.

Note that in our example for all $t, w > 0$

$$\left|\Pr\left(s_{t+w}^{-i} = s^{-i} \mid h_{t+w-1}\right) - \Pr\left(s_{t+w}^{-i} = s^{-i} \mid h'_{t+w-1}\right)\right| = \begin{cases} 0 & \text{if } w \neq M \\ \rho & \text{if } w = M \end{cases}$$

for any two histories h_{t+w-1} and h'_{t+w-1} such that for every $\tau < t + w$, $\tau \neq t$ we have $s_\tau = s'_\tau$ and for $\tau = t$ we have $s_\tau^{-i} = s_\tau^{-i}$, $s_\tau^i \neq s_\tau^i$ (that is, the two histories h_{t+w-1} and h'_{t+w-1} differ only in i's action at time t).

However, $O\left(\frac{f(M)}{g(t)}\right) \to 0$ for any g such that $g \to \infty$; hence our condition (3.1) is not fulfilled.

Claim: *The regrets of player* 1 *do not converge to zero with probability* 1.

Proof: Suppose that the regrets of player 1 do converge to zero with probability 1; we shall show that this leads to a contradiction.

For all positive integers $t, k > 0$ we have

$$\Pr\left[s_{t+k+M}^2 = R \mid s_{t+M}^2 = R \text{ and } s_{t+i}^1 = T \text{ for } i = 1, ..., k\right] = 1,$$

and

$$\Pr\left[s_{t+k+M}^2 = R \mid s_{t+M}^2 = L \text{ and } s_{t+i}^1 = T \text{ for } i = 1, ..., k\right] = 1 - (1-\rho)^k$$

(since, once a switch from L to R occurs, s^2 remains at R; the probability of no switch, thus always L, is $(1-\rho)^k$). Hence

$$\Pr\left[s_{t+k+M}^2 = R \mid s_{t+i}^1 = T \text{ for } i = 1, ..., k\right] \geq 1 - (1-\rho)^k. \tag{3.3}$$

The same argument implies that for any positive integer $c > k$

$$\Pr[s^2_{t+M+j} = R \text{ for } j = k, ..., c \mid s^1_{t+i} = T \text{ for } i = 1, ..., c] \geq 1 - (1-\rho)^k \,, \quad (3.4)$$

since once player 2 is at R he will not switch to L. The same argument obviously holds for the pair of actions (B, L), replacing T by B and R by L; hence

$$\Pr[s^2_{t+M+j} = L \text{ for } j = k, ..., c \mid s^1_{t+i} = B \text{ for } i = 1, ..., c] \geq 1 - (1-\rho)^k \,. \quad (3.5)$$

Let $\varepsilon > 0$, let k be such that $(1-\rho)^k < \varepsilon$, and let c be such that $2\,(M+k)\,/c < \varepsilon$. We divide time into blocks of length c. Let

$$H_1 := \{1, 2, ..., c\}\,, \, H_2 := \{c+1, c+2, ..., 2c\}\,, ... \,.$$

Let

$$X_i := \frac{1}{c} \sum_{v \in H_i : s^1_v = T} \left[u^1\left(B, s^2_v\right) - u^1\left(s^1_v, s^2_v\right)\right]\,, \qquad i = 1, 2, 3, ...$$

(note that $|X_i| \leq 1$).[9] Similarly, define

$$Y_i := \frac{1}{c} \sum_{v \in H_i : s^1_v = B} \left[u^1\left(T, s^2_v\right) - u^1\left(s^1_v, s^2_v\right)\right]\,, \qquad i = 1, 2, 3, ... \,.$$

Suppose no change has occurred in player 1's play in block H_i (denote this event by $\widetilde{H_i}$); then by (3.4) and (3.5) we get

$$\mathrm{E}\left[X_i + Y_i \mid \widetilde{H_i}\right] > (1-\varepsilon)^2 - \varepsilon > 1 - 3\varepsilon. \quad (3.6)$$

Indeed, if the action of player 1 had been T (denote this event by $\widetilde{H_{i,T}}$), then by (3.4) $\Pr[s^2_{t+M+j} = R \text{ for } j = k, ..., c] \geq 1 - (1-\rho)^k > 1 - \varepsilon$. Therefore

$$\begin{aligned} \mathrm{E}\left[X_i + Y_i \mid \widetilde{H_{i,T}}\right] &> \left(\frac{c-(M+k)}{c} - \frac{M+k}{c}\right) \Pr[s^2_{t+M+j} = R \text{ for } j = k, ..., c \mid \widetilde{H_{i,T}}] \\ &\quad - \left(1 - \Pr[s^2_{t+M+j} = R \text{ for } j = k, ..., c \mid \widetilde{H_{i,T}}]\right) \\ &> (1-\varepsilon)^2 - \varepsilon > 1 - 3\varepsilon. \end{aligned}$$

We can now check the frequency of no change in player 1's action. The probability of a change from t to $t+1$ is either $(1/\mu)\, R^1_t\,(T, B)$ or $(1/\mu)\, R^1_t\,(B, T)$,

[9] Since there are only two strategies for player 1 we can conclude that $X_i = (1/c) \sum_{v \in H^1_i}^{c} \left[u\left(B, s^2_{ti+v}\right) - u^1\left(s^1_{ti+v}, s^2_{ti+v}\right)\right]$, which is the ith block Hannan [1957] regret, which in this case coincides with the Hart–Mas-Colell regret.

hence less than $(1/\mu)\, R_t^1\,(T,B) + (1/\mu)\, R_t^1\,(B,T)$. Therefore in a block of length c the probability of a change is no more than

$$c \cdot \frac{1}{\mu} \max_{t \in H_i} \left\{ R_t^1\,(T,B) + R_t^1\,(B,T) \right\}.$$

Now since $R_t^1\,(T,B)$ and $R_t^1\,(B,T)$ converge to zero a.s. there exists i_0 such that

$$\Pr\left[\frac{1}{\mu} \max_{t \in H_i} \left\{ R_t^1\,(T,B) + R_t^1\,(B,T) \right\} < \frac{\varepsilon}{c} \right] > 1 - \varepsilon$$

for all $i > i_0$. Hence the probability of a change in player 1's action in block H_i is less than $c \cdot \varepsilon / c \cdot (1-\varepsilon) + \varepsilon < 2\varepsilon$. Therefore

$$\begin{aligned} \mathrm{E}\left[X_i + Y_i\right] &\geq \mathrm{E}\left[X_i + Y_i \mid \widetilde{H_i}\right] \Pr\left(\widetilde{H_i}\right) - \left(1 - \Pr\left(\widetilde{H_i}\right)\right) \\ &> (1-2\varepsilon)(1-3\varepsilon) - 2\varepsilon > 1 - 7\varepsilon \end{aligned}$$

for all $i > i_0$. We can easily choose ε such that $1 - 7\varepsilon > 0.9$; hence the average satisfies

$$\liminf_{n \to \infty} \mathrm{E}\left[\overline{X_n + Y_n}\right] > 0.9.$$

But $\liminf_{n\to\infty} \mathrm{E}\left[\overline{X_n + Y_n}\right]$ is less than or equal to $\lim_{t\to\infty} R_t^1\,(T,B) + R_t^1\,(B,T)$, which contradicts our assumption. ■

- An interesting question is, what happens if our $g\,(t)$ does not converge to infinity but has a subsequence which converges to infinity. Can one still get convergence of the regrets to zero? (The answer probably depends on how dense this subsequence is.)

- Another interesting question is, what happens if we do not have f,g, as in (3.1) but only require that for any w:

  $$\left|\Pr\left(s^{-i} \mid h_{t+w-1}\right) - \Pr\left(s^{-i} \mid h_t, s_{t+1}^{-i}, ..., s_{t+w-1}^{-i}\right)\right| \to_{t\to\infty} 0.$$

  Can one still get convergence of the regrets to zero?

4. A General Class of Simple Adaptive Procedures

4.1. Introduction

As explained in Section 2.2 above, in [HM2] Hart and Mas-Colell exhibit a class of adaptive strategies that have a convergence property. Specifically, they define a function Λ defined on all $\mathbb{R}^m$ except for a closed and convex approachable set C, which defines "directions" from this set C. Using a strategy which follows Λ in some sense causes the average payoffs to converge to the set C. Applying this to the setup of conditional regrets — see [HM2, Section 5.1] — yields strategies that require the computation of an eigenvector at every step. The question raised in [HM2] is whether we can find a simple adaptive procedure with reference to Λ, in the same way as is done [HM1] (which corresponds to the special case of the l_2-potential). In this section we show that the answer to this question is in the affirmative.

4.2. The Model

Consider a game in strategic form played by a finite set of players N, each having a finite set of strategies S^i. Fix a player i, let $m^i := \left|S^i\right|$ and $L^i := \left\{(j,k) : j \neq k \text{ and } j,k \in S^i\right\}$. (From now on we omit i whenever it is obvious that we are dealing with player i.) Let K' be a closed bounded set in $\mathbb{R}^L$ containing in its interior the set of all possible payoffs of player i; we can without loss of generality assume that $0 \in K'$ (as in Section 2.2, we view the vector of regrets of player i in the one-shot game as his payoff vector). Let $K := K' + (-K')$; note that K is compact. Let $C \subseteq \mathbb{R}^L$ be a closed convex set[10] such that $C \supseteq \mathbb{R}^L_-$. Let $w : \mathbb{R}^L \to \mathbb{R}$ be defined by $w(x) := \sup_{y \in C} \{x \cdot y\}$. Notice that since $C \supseteq \mathbb{R}^L_-$,

$$w(x) \geq 0 \text{ if } x(j,k) \geq 0 \text{ for all } j \neq k, \text{ and } w(x) = \infty \text{ otherwise.} \tag{4.1}$$

Let Λ and P be as in [HM2, Section 2]; i.e., $\Lambda : \mathbb{R}^L \backslash C \to \mathbb{R}^L$, and $P : \mathbb{R}^L \to \mathbb{R}$ but with the following slightly stronger conditions:[11]

(D1) Λ is Lipschitz on $K \backslash C$.

(D2) P is differentiable; ∇P is Lipschitz on K; and $\nabla P(x) = \phi(x)\Lambda(x)$ for almost every $x \notin C$, where $\phi : \mathbb{R}^L \backslash C \to \mathbb{R}_{++}$ is a continuous positive function.

(D3) $\Lambda(x) \cdot x > w(\Lambda(x))$ for all $x \notin C$.

(D4) Λ can be extended to a Lipschitz function on[12] K.

[10] Using a general set C, rather than $\mathbb{R}^L_-$, allows us to handle strategies like Fudenberg and Levine's [1995, 1998, 1999] smooth fictitious play. This will be discussed later in Section 4.5.

[11] The change is that in both (D1) and (D2) we require Λ and ∇P to be Lipschitz rather than just continuous, and we have added (D4).

[12] This added condition is not very strong. In most cases we have a trivial extension to Λ,

Notice that (D1) and (D2) imply $\nabla P(x) = \phi(x)\Lambda(x)$ for every $x \notin C$. By (4.1), (D2), and (D3), for every $x \notin C$ we have $\Lambda(x)(j,k) \geq 0$ and $\nabla P(x)(j,k) \geq 0$ for all $j \neq k$; therefore we can w.l.o.g. also assume in (D4) $\Lambda(x)(j,k) \geq 0$ on[13] K.

Define $a_t^i, d_t^i, \lambda_t^i \in \mathbb{R}^L$ by

$$\begin{aligned} a_t^i(j,k) &:= \begin{cases} 0 & \text{if } s_t^i \neq j \\ u^i\left(k, s_t^{-i}\right) - u^i\left(j, s_t^{-i}\right) & \text{if } s_t^i = j \end{cases} \\ d_t^i &:= \frac{1}{t}\sum_{v=1}^{t} a_v^i \\ \lambda_t^i &:= \Lambda\left(d_t^i\right). \end{aligned}$$

Notice that $\lambda_t^i(j,k) \geq 0$ for all j,k. Since a_t^i, d_t^i lie in the compact set K^L and Λ is continuous, it follows that λ_t^i is bounded.

Let $\mu > 0$ be large enough for the following π_t^i to be a probability function such that[14] $\pi_t^i(j,j) > 0$. For every $j \in S^i$, let

$$\pi_t^i(j,k) := \frac{1}{\mu}\lambda_t^i(j,k) \text{ for } k \neq j \text{ and } \pi_t^i(j,j) := 1 - \sum_{k \in S^i : k \neq j} \pi_t^i(j,k) \tag{4.2}$$

be the transition probabilities from stage t to $t+1$ (we can let the probabilities π_0^i of the first move s_1^i be arbitrary). Notice that these probabilities, the set C, and the functions Λ and P are defined separately for every different player; for convenience, we drop the superscript i when it is clear.

Now assume that for every player i' there is a set $C^{i'}$ and functions $P^{i'}$ and $\Lambda^{i'}$. We show that if all players use these strategies (which are similar to HMS) as in [HM1], then the regrets of every player i' will converge to his set $C^{i'}$. Furthermore, as in Section 3, we show that even if the other players do not follow this strategy, but change their actions with only slight connection to player i's actions, then player i's regrets converge to C^i.

Theorem 4.1. *If every player i uses the strategy given in (4.2), then $d_t^i \to_{t\to\infty} C^i$ a.s. for every player i.*

since ∇P is proportional to Λ (notice that ∇P is Lipschitz on K). This condition is needed since, unlike [HM2], we want a simple procedure in which probabilities are proportional to Λ; therefore Λ should be defined globally.

[13] One can take $\Lambda^*(j,k) := \max\{0, \Lambda(j,k)\}$ instead of Λ.

[14] Any μ greater than $\max_{x \in K} \sum_{k \neq j} \Lambda_{(k,j)}(x)$ will suffice.

Suppose now that $-i$ does not follow the above procedure but, as in the previous section, $-i$ plays in a way that given two histories h_t and h'_t such that for every $\tau \neq t-w$ we have $s_\tau = s'_\tau$ and for $\tau = t-w$ we have $s^{-i}_\tau = s'^{-i}_\tau$ and possibly $s^i_\tau \neq s'^i_\tau$, then

$$\left|\Pr\left(s^{-i}_{t+1} = s^{-i} \mid h_t\right) - \Pr\left(s^{-i}_{t+1} = s^{-i} \mid h'_t\right)\right| \leq \frac{f(w)}{g(t)} \tag{4.3}$$

for some functions f, g such that $g(t) \rightarrow_{t\rightarrow\infty} \infty$.

Theorem 4.2. *If player i uses strategy (4.2), i.e., $\Pr\left(s^i_{t+1} = s^i \mid h_t\right) = \pi^i_t\left(s^i_t, s^i\right)$, then it is guaranteed that $d^i_t \rightarrow_{t\rightarrow\infty} C^i$ a.s. for any strategies of the other players satisfying (4.3).*

4.3. Proof of Theorem 4.1

Lemma 2.3 of [HM2] shows that there exists a constant c such that

$$\begin{array}{ll} P(x) = c & \text{if } x \in \partial C \ (= \text{the boundary of } C) \\ P(x) > c & \text{if } x \notin C. \end{array}$$

Choose $\varepsilon > 0$. We shall show that $\Pr(\limsup P(d_t) \leq c + \varepsilon) = 1$, and since this is true for every ε, it follows that $\Pr(\limsup P(d_t) \leq c) = 1$, which is equivalent to $d_t \rightarrow_{t\rightarrow\infty} C$ a.s., since P is continuous.

Let P_1 be as in [HM2, Section 2.2], namely $P_1(x) := [P(x) - c]^2$ for every $x \notin C$ and $P_1(x) := 0$ for $x \in C$. Notice that

$$\nabla P_1(x) = 2\nabla P(x)\,[P(x) - c]\,.$$

Let $Q : \mathbb{R}^L \rightarrow \mathbb{R}$ be as in [HM2, Section 2.2, Proof of Theorem 2.1], i.e.,

$$Q(x) \geq 0 \text{ for all } x \in \mathbb{R}^L, \text{ and } Q(x) = 0 \text{ if and only if } P_1(x) \leq \varepsilon; \tag{4.4}$$

$$\nabla Q(x) \cdot x - w(\nabla Q(x)) \geq Q(x)\,; \text{ and} \tag{4.5}$$

$$\nabla Q(x) = \begin{cases} 0, & \text{if } \ Q(x) = 0, \\ r\,(P_1(x) - \varepsilon)^{r-1}\, 2\nabla P(x)\,[P(x) - c], & \text{otherwise.} \end{cases} \tag{4.6}$$

Let $y(x) := r\left[P_1(x) - \varepsilon\right]_+^{r-1} 2\left[P(x) - c\right]$; now by (4.4) we can write

$$\nabla Q(x) = y(x) \cdot \nabla P(x). \tag{4.7}$$

Notice that by (4.1) it follows that (4.5) can be written as:

$$\nabla Q(x)(j,k) \geq 0 \text{ for all } j \neq k, \text{ and } \nabla Q(x) \cdot x \geq Q(x). \tag{4.8}$$

Let $q_t := \nabla Q(d_t)$. Notice that by (D2) and the fact that $q_t = 0, \lambda_t = 0$ for all $x \in C$ it follows that

$$q_t = \lambda_t \cdot y(d_t). \tag{4.9}$$

In the sequel, the present proof will be divided into steps similar to those of the Proof of the Main Theorem in [HM1, Appendix].

4.3.1. Steps of the Proof

- *Step N1:*

(i) $\mathrm{E}\left[(t+v) Q(d_{t+v}) \mid h_t\right] \leq tQ(d_t) + \sum_{w=1}^{v} \mathrm{E}\left[a_{t+w} \mid h_t\right] \cdot q_t + O\left(\frac{v^2}{t}\right).$
(ii) $(t+v) Q(d_{t+v}) - tQ(d_t) = O(v).$

Proof:

$$\begin{aligned} Q(d_{t+v}) &= Q\left(\frac{t}{t+v} d_t + \frac{1}{t+v}\sum_{w=1}^{v} a_{t+w}\right) \\ &= Q(d_t) + \left(\frac{t}{t+v} d_t + \frac{1}{t+v}\sum_{w=1}^{v} a_{t+w} - d_t\right) \cdot \nabla Q(d_t) \\ &\quad + O\left(\frac{t}{t+v} d_t + \frac{1}{t+v}\sum_{w=1}^{v} a_{t+w} - d_t\right)^2. \end{aligned}$$

The second equality follows since ∇Q is Lipschitz[15] on K and hence there exists $0 \leq t \leq 1$ such that

$$\begin{aligned} Q(x+y) &= Q(x) + y \cdot \nabla Q(x+ty) \\ &\leq Q(x) + y \cdot \nabla Q(x) + \|y\| \, k \, \|ty\| = Q(x) + y \cdot \nabla Q(x) + O\left(\|y\|^2\right). \end{aligned}$$

[15] The Lipschitz constant depends on ε, but this does not affect the proof.

By (4.8) $\nabla Q(d_t) \cdot d_t \geq Q(d_t)$, hence

$$\begin{aligned}(t+v)\, Q(d_{t+v}) &= (t+v)\, Q(d_t) + \left(\sum_{w=1}^{v} (a_{t+w} - d_t)\right) \cdot \nabla Q(d_t) + O\left(\frac{v^2}{(t+v)}\right) \\ &\leq tQ(d_t) + \sum_{w=1}^{v} a_{t+w} \cdot q_t + O\left(\frac{v^2}{t+v}\right).\end{aligned}$$

Taking expectation yields (i), and since $O\left(\frac{v^2}{(t+v)}\right) \leq O(v)$ and a_{t+w},q_t are bounded we get (ii). □

Define

$$\alpha_{t,w}\left(j, s^{-i}\right) := \sum_{k \in S^i} \pi_t^i(j,k) \Pr\left[s_{t+w} = \left(k, s^{-i}\right) \mid h_t\right] - \Pr\left[s_{t+w} = \left(j, s^{-i}\right) \mid h_t\right].$$

- *Step N2:*

$$\mathrm{E}\left[a_{t+w} \mid h_t\right] \cdot q_t = \mu y(d_t) \cdot \sum_{(j,s^{-i}) \in S^i \times S^{-i}} \alpha_{t,w}\left(j, s^{-i}\right) u^i\left(j, s_t^{-i}\right).$$

Proof: By the same method as in Step *M2* of [HM1] we get:

$$\mathrm{E}\left[a_{t+w} \mid h_t\right] \cdot \lambda_t = \sum_{(j,s^{-i}) \in S^i \times S^{-i}} \alpha_{t,w}\left(j, s^{-i}\right) u^i\left(j, s_t^{-i}\right)$$

and (4.9) yields the result. □

- *Step N3:*

(i) $d_{t+v}(j,k) - d_t(j,k) = O\left(\frac{v}{t}\right)$.
(ii) $q_{t+v}(j,k) - q_t(j,k) = O\left(\frac{v}{t}\right)$.
(iii) $\lambda_{t+v}(j,k) - \lambda_t(j,k) = O\left(\frac{v}{t}\right)$.
(iv) $\pi_{t+v}(j,k) - \pi_t(j,k) = O\left(\frac{v}{t}\right)$.

Proof: Since $d_{t+v} = d_t + \frac{1}{t+v} \sum_{w=1}^{v} (a_{t+w} - d_t)$ and $(a_{t+w} - d_t)$ is bounded, (i) is true. Since by (4.6) ∇Q is Lipschitz on the compact set K, (ii) follows. The fact that Λ is Lipschitz yields (iii) and (iv). □

- *Steps N4–N7* are exactly the same as Steps M4–M7 in [HM1] based on *N3(iv)*.

- *Step N8:* $\mathrm{E}\left[(t+v)\, Q(d_{t+v}) \mid h_t\right] \leq tQ(d_t) + O\left(\frac{v^3}{t} + v^{\frac{1}{2}}\right)$.

Proof: Steps *N5* and *N7* imply that $\alpha_{t,w}\left(j,s^{-i}\right)=O\left(\frac{w^2}{t}+w^{-0.5}\right)$. The fact that K is compact and that y is continuous yields that y is bounded on K. Therefore the formula of Step *N2* yields $\mathrm{E}\left[a_{t+w}\mid h_t\right]\cdot q_t=O\left(\frac{w^2}{t}+w^{-0.5}\right)$. Summing over w and Step *N1(i)* yield the result. □

Let $t_n:=\left\lfloor n^{5/3}\right\rfloor$, and let $v_n:=t_{n+1}-t_n=O\left(n^{2/3}\right)$.

- *Step N9:* $\mathrm{E}\left[t_{n+1}Q\left(d_{t_{n+1}}\right)\mid h_{t_n}\right]\leq t_nQ\left(d_{t_n}\right)+O\left(n^{1/3}\right)$.

Proof: Immediate by Step *N8*. □

- *Step N10*: $\lim_{n\to\infty}Q\left(d_{t_n}\right)\to 0$ a.s.

Proof: Define $b_n:=t_n\approx n^{5/3}$ and $X_n:=b_nQ\left(d_{t_n}\right)-b_{n-1}Q\left(d_{t_{n-1}}\right)$.

By Step *N1(ii)* we have $|X_n|\leq O\left(v_n\right)=O\left(t_n/n\right)$, thus

$$\sum_n\frac{1}{b_n^2}\operatorname{Var}\left(X_n\right)=\sum_nO\left(\frac{1}{n^2}\right)<\infty.$$

Next, by Step *N9* we have

$$\frac{1}{b_n}\sum_{\nu\leq n}\mathrm{E}\left[X_\nu\mid X_1,...,X_{\nu-1}\right]\leq O\left(n^{-5/3}\sum_{\nu\leq n}\nu^{1/3}\right)=O\left(n^{-1/3}\right)\to 0.$$

Applying the Strong Law of Large Numbers for Dependent Random Variables yields

$$\frac{1}{b_n}\sum_{\nu\leq n}X_\nu=Q\left(d_{t_n}\right)\to_{n\to\infty}0$$

a.s. □

- *Step N11:*

(i) $\lim_{t\to\infty}Q\left(d_t\right)\to 0$ a.s.

(ii) $\Pr\left(\limsup P\left(d_t\right)\leq c+\varepsilon\right)=1$.

Proof: Since for $t_n<t\leq t_{n+1}$ we have $\frac{t-t_n}{t_n}\leq\frac{v_n}{t_n}=O\left(n^{-1}\right)$, by *N3* we get $Q\left(d_t\right)\to_{t\to\infty}0$ a.s., and by (4.4) and the fact that Q and P are continuous it follows that $\Pr\left(\limsup P\left(d_t\right)\leq c+\varepsilon\right)=1$. ■

4.4. Proof of Theorem 4.2

It is easy to see that all the claims in Section 3.3.1 about f, g and the histories h, h' hold also in this case. The Proof of Theorem 4.1 can be used for the Proof of Theorem 4.2; the only changes necessary are precisely those that were needed in the Proof of Theorem 3.1 of Section 3. (We use lowercase letters for the steps in this proof.) Namely, in Step *n4* define $\hat{s}_{t+w}$ differently. Here $\hat{s}_{t+w}$ will be defined by $\hat{s}_t := s_t$, and the transition probabilities are:

$$\Pr\left(\hat{s}_{t+w} = s \mid \hat{s}_t, ..., \hat{s}_{t+w-1}\right) = \pi_t^i\left(\hat{s}_{t+w-1}^i, s^i\right) \cdot \Pr\left(s_{t+w}^{-i} = s^{-i} \mid h_t, \hat{s}_{t+1}^{-i}, ..., \hat{s}_{t+w-1}^{-i}\right)$$

and we get results similar to *m4*-*m7* of Section 3. In Step *n8* we get

$$\mathrm{E}\left[(t+v)\, Q\left(d_{t+v}\right) \mid h_t\right] \leq tQ\left(d_t\right) + O\left(\frac{f(v)\, v^2}{g(t)} + v^{\frac{1}{2}}\right),$$

similar to *m8*.

We now need Steps *n9.1* and *n9.2* similar to *m9.1* and *m9.2*. Specifically:

- *Step n9.1:* $E\left[t_{n+1}Q\left(d_{t_{n+1}}\right) \mid h_{t_n}\right] \leq t_n Q\left(d_{t_n}\right) + O\left(\frac{v_n^2 f(v_n)}{g(t_n)} + v_n^{1/2}\right)$.

  This follows immediately from *n8*.

Define $\widetilde{f}(w) := w^2 f(w)$. (Notice that $\widetilde{f}$ has an inverse function, denoted by $\widetilde{f}^{-1}$.) Let $a_n := \frac{1}{2}\widetilde{f}^{-1}(g(n))$, and let $t_n := \lceil n a_n \rceil$; $(v_n = t_{n+1} - t_n)$.

- *Step n9.2:*

(i) a_n is a nondecreasing sequence, and $a_n \rightarrow_{n\rightarrow\infty} \infty$.
(ii) $\frac{v_n}{t_n} = O\left(n^{-1}\right)$.
(iii) $\widetilde{f}(v_n)/g(t_n) = O(1)$.
(iv) $\frac{v_n^2 f(v_n)}{g(t_n)} + v_n^{1/2} = O\left(a_n^{1/2}\right)$.

The only thing different from Step *m9.2* is *(iv),* which follows immediately from *(iii)* and the fact that $v = O(a_n)$ (proved in Step *m9.2*).

The rest of the proof is the same[16] as in Theorem 4.1. ■

[16] In Step *n10* we get (with $b_n = t_n$)

$$\sum_n \mathrm{Var}\left(X_n\right)/b_n^2 = \sum_n O\left(n^{-2}\right) < \infty$$

4.5. Conditional Smooth Fictitious Play

Fictitious play is a strategy where a player plays a best reply to the empirical distribution of play z_t. It may be viewed as a regret-based strategy, corresponding to the l_∞-potential (cf. [HM2, Section 4.1]). However, this strategy does not satisfy the above conditions, specifically condition (D1): Λ is not continuous. Fudenberg and Levine [1995, 1998] present a smoothing of fictitious play. Let $\sigma^i_{t+1} \in \Delta\left(S^i\right)$ denote the (possibly mixed) choice of player i at time $t+1$. Fictitious play requires

$$\sigma^i_{t+1} \in \underset{\sigma^i \in \Delta(S^i)}{\operatorname{argmax}} \left\{ u^i\left(\sigma^i, z^{-i}_t\right)\right\}$$

(note that the maximizer is not necessarily unique). In *smooth fictitious play*, we have instead

$$\sigma^i_{t+1} = \underset{\sigma^i \in \Delta(S^i)}{\operatorname{argmax}} \left\{ u^i\left(\sigma^i, z^{-i}_t\right) + \nu\left(\sigma^i\right)\right\}$$

where $\nu \equiv \nu^i$ is a smooth strictly differentiably concave function with gradient vector approaching infinite length as one approaches the boundary of $\Delta(S^i)$; hence, the maximizer is unique.

In the conditional case (Fudenberg and Levine [1998, 1999]), instead of z^{-i}_t one considers for every $j \in S^i$ the distribution of play $z^i_t(j)$ only in those periods where i played j, i.e.,

$$z^{-i}_t(j)(s^{-i}) := \frac{1}{t}\left|\left\{\tau \le t : s_\tau = (j, s^{-i})\right\}\right|.$$

Since u^i is linear,

$$\underset{\sigma^i \in \Delta(S^i)}{\operatorname{argmax}} \left\{ u^i\left(\sigma^i, z^{-i}_t(j)\right) + \nu\left(\sigma^i\right)\right\} = \underset{\sigma^i \in \Delta(S^i)}{\operatorname{argmax}} \left\{ \sigma^i \cdot D^i_t(j, \cdot) + \nu\left(\sigma^i\right)\right\}$$

by Steps *n1(ii)* and *n9.2(ii)*. The following inequality

$$\frac{1}{b_n}\sum_{\nu \le n} \mathrm{E}\left[X_\nu \mid X_1, ..., X_{\nu-1}\right] \le O\left(n^{-1}a_n^{-1}\sum_{\nu \le n} a_\nu^{0.5}\right)$$

we get by Steps *n9.1* and *n9.2(iv)*, and since a_n is increasing

$$\sum_{\nu \le n} a_\nu^{0.5} \le n a_n^{0.5};$$

hence

$$\frac{1}{b_n}\sum_{\nu \le n} \mathrm{E}\left[X_\nu \mid X_1, ..., X_{\nu-1}\right] \to 0.$$

(we write $\sigma^i \cdot D_t^i(j,\cdot)$ for $\sum_{k\neq j} \sigma^i(k) \cdot D_t^i(j,k)$).

Therefore we define "*conditional smooth fictitious play*" as the Hart–Mas-Colell conditional-regret-based strategy (4.2), with $\Lambda(x) := \nabla P(x)$, where

$$P(x) := \sum_{j \in S^i} \max_{\sigma_j^i \in \Delta(S^i)} \left\{ \sum_{k \in S^i, k\neq j} \sigma_j^i(k) \cdot x(j,k) + \nu(\sigma_j^i) \right\}, \text{ for all } x \in \mathbb{R}^L \quad (4.10)$$

is the corresponding potential. It follows immediately from the definition of a Λ-strategy in [HM2, Section 2.1] that the argmax is a Λ-strategy. Notice that $\operatorname{argmax}_{\sigma_j^i \in \Delta(S^i)} \left\{ \sigma_j^i \cdot x(j,\cdot) + \nu(\sigma_j^i) \right\}$ is a smooth (and in particular twice differentiable) function, as shown by Fudenberg and Levine [1999, Section 3]; thus P and Λ satisfy properties (D1)–(D4) for[17] $C := \{x \mid P(x) \leq m \|\nu\|\}$ (again, see [HM2, Section 4.1]).

Now, by our results we get that if player i plays conditional smooth fictitious play as defined above, and for the other players (4.3) holds, then all the conditional regrets of player i will in the limit be at most $m\|\nu\|$ (a.s.). Formally, this can be written, according to Fudenberg and Levine's notations, as:

Proposition 4.3. *The strategy (4.2), where π_t^i are transition probabilities given by the conditional smooth fictitious play potential (4.10), is $m^i \left\|\nu^i\right\|$-calibrated for any strategies of the other players satisfying (4.3). Moreover, if all players play this way, then the empirical distribution of play z_t converges a.s. to the set of correlated ε-equilibria, where $\varepsilon = \max_{i\in N} m^i \left\|\nu^i\right\|$.*

The difference between our strategy and that of Fudenberg and Levine is that we do not have to evaluate eigenvectors, as do Fudenberg and Levine, but our probabilities are just proportional to Λ.

References

Aumann, R. J. [1974], Subjectivity and Correlation in Randomized Strategies, *Journal of Mathematical Economics* 1, 67–96.

Blackwell, D. [1956], An Analog of the Minmax Theorem for Vector Payoffs, *Pacific Journal of Mathematics* 6, 1–8.

[17] Recall that $m \equiv m^i := |S^i|$.

Fudenberg, D. and D. K. Levine [1995], Universal Consistency and Cautious Fictitious Play, *Journal of Economic Dynamics and Control* 19, 1065–1090.

Fudenberg, D. and D. K. Levine [1998], *Theory of Learning in Games,* MIT Press.

Fudenberg, D. and D. K. Levine [1999], Conditional Universal Consistency, *Games and Economic Behavior* 29, 104–130.

Hannan, J. [1957], Approximation to Bayes Risk in Repeated Play, in *Contributions to the Theory of Games, Vol. III* (*Annals of Mathematics Studies 39*), M. Dresher, A. W. Tucker and P. Wolfe (eds.), Princeton University Press, 97–139.

Hart, S. and A. Mas-Colell [2000], A Simple Adaptive Procedure Leading to Correlated Equilibrium, *Econometrica* 68, 1127–1150. **[HM1]**

Hart, S. and A. Mas-Colell [2001], A General Class of Adaptive Strategies, *Journal of Economic Theory* 98, 26–54. **[HM2]**

Loève, M. [1978], *Probability Theory, Vol. II*, 4th Edition, Springer-Verlag.