Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Games and Economic Behavior 45 (2003) 375–394
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Regret-based continuous-time dynamics ☆

Sergiu Hart [a,*] and Andreu Mas-Colell [b]

[a] *Center for Rationality and Interactive Decision Theory, Department of Mathematics, and Department of Economics, The Hebrew University of Jerusalem, Feldman Building, Givat Ram, 91904 Jerusalem, Israel*
[b] *Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain*

Received 8 January 2003

Abstract

Regret-based dynamics have been introduced and studied in the context of discrete-time repeated play. Here we carry out the corresponding analysis in continuous time. We observe that, in contrast to (smooth) fictitious play or to evolutionary models, the appropriate state space for this analysis is the space of distributions on the product of the players' pure action spaces (rather than the product of their mixed action spaces). We obtain relatively simple proofs for some results known in the discrete case (related to 'no-regret' and correlated equilibria), and also a new result on two-person potential games (for this result we also provide a discrete-time proof).

JEL classification: C7; D7; C6

1. Introduction

'Regret-matching' as a strategy of play in long-run interactions has been introduced and studied in a number of earlier papers (Hart and Mas-Colell, 2000, 2001a, 2001b). We have shown that, under general conditions, regret-matching leads to distributions of play that are related to the concept of correlated equilibrium. The purpose of the current paper is to reexamine the dynamics of regret-matching from the standpoint of differential dynamics in continuous time. It is well known that this approach often leads to a simplified and

☆ Previous version: August 2001.
* Corresponding author.
E-mail addresses: hart@huji.ac.il (S. Hart), mcolell@upf.es (A. Mas-Colell).
URL: http://www.ma.huji.ac.il/~hart.

0899-8256/$ – see front matter
doi:10.1016/S0899-8256(03)00178-7

streamlined treatment of the dynamics, to new insights and also to new results—and this will indeed happen here.

An important insight comes already in the task of formulating the differential setup. The appropriate state space for regret-matching is not the product of the mixed action spaces of the players but a larger set: the distributions on the product of the pure action spaces of the players. Of course the players play independently at every point in time—but this in no way implies that the state variable evolves over time as a product distribution.

In Section 2 we present the model and specify the general setup of the dynamics we consider. In Section 3 we analyze general regret-based dynamics, the continuous-time analog to Hart and Mas-Colell (2001a), to which we refer for extensive discussion and motivation. In Section 4 we establish that for some particularly well-behaved classes of two-person games—zero-sum games, and potential games—the dynamics in fact single out the Nash equilibria of the game. The result for potential games is new and so we present a discrete-time version in Appendix A. In Section 5 we move to the analysis of conditional regret dynamics and prove convergence to the set of correlated equilibria. Finally, Section 6 offers some remarks. Appendix B provides a technical result, and Appendix C deals with the continuous-time version of the approachability theorems à la Blackwell (1956), which are basic mathematical tools for this area of research.

We dedicate this paper to the memory of Bob Rosenthal. It is not really necessary to justify this by exhibiting a connection between our topics of interest here and some particular paper of his. The broadness of his intellectual gaze guarantees that he would have been engaged and that, as usual, he would have contributed the insightful comments that were a trademark of his. At any rate, we mention that one of the cases that we examine with some attention is that of potential games and that Bob Rosenthal was the first to identify the remarkable properties of this class of games (Rosenthal, 1973).

2. Model

2.1. Preliminaries

An *N-person game Γ in strategic form* is given by a finite set N of *players*, and, for each player $i \in N$, by a finite set S^i of *actions* and a *payoff function* $u^i : S \to \mathbb{R}$, where $S := \prod_{i \in N} S^i$ is the set of N-tuples of actions (we call the elements of S^i 'actions' rather than strategies, a term we will use for the repeated game). We write $S^{-i} := \prod_{j \in N,\ j \neq i} S^j$ for the set of action profiles of all players except player i, and also $s = (s^i, s^{-i})$. Let M be a bound on payoffs: $|u^i(s)| \leqslant M$ for all $i \in N$ and all $s \in S$.

A randomized (mixed) action x^i of player i is a probability distribution over i's pure actions, i.e.,[1] $x^i \in \Delta(S^i)$. A *randomized joint action* (or *joint distribution*) z is a probability distribution over the set of N-tuples of pure actions S, i.e., $z \in \Delta(S)$. Given such z, we write $z^i \in \Delta(S^i)$ and $z^{-i} \in \Delta(S^{-i})$ for the marginals of z, i.e., $z^i(s^i) = \sum_{s^{-i} \in S^{-i}} z(s^i, s^{-i})$

[1] For a finite set A, we write $|A|$ for the number of elements of A, and $\Delta(A)$ for the set of probability distributions on A, i.e., $\Delta(A) := \{x \in \mathbb{R}^A_+ \colon \sum_{a \in A} x(a) = 1\}$ (the $(|A| - 1)$-dimensional unit simplex).

for all $s^i \in S^i$, and $z^{-i}(s^{-i}) = \sum_{s^i \in S^i} z(s^i, s^{-i})$ for all $s^{-i} \in S^{-i}$. When the joint action is the result of independent randomizations by the players, we have $z(s) = \prod_{i \in N} z^i(s^i)$ for all $s \in S$; we will say in this case that z is *independent*, or that it is a *product measure*.[2]

2.2. Dynamics

We consider continuous-time dynamics on $\Delta(S)$ of the form

$$\dot{z}(t) = \frac{1}{t}\big(q(t) - z(t)\big), \tag{2.1}$$

where $q(t) \in \Delta(S)$ is[3] the joint play at time t and $z(t)$ is the 'time-average joint play.' Assume one starts at $t = 1$ with some[4] $z(1) \in \Delta(S)$.

To justify (2.1), recall the discrete-time model: Time is $t = 1, 2, \ldots$; player i at period t plays $s_t^i \in S^i$, and the time-average joint play at time t is $z_t \in \Delta(S)$, given inductively by[5] $z_t = (1/t)(\mathbf{1}_{s_t} + (t-1)z_{t-1})$, or

$$z_t - z_{t-1} = \frac{1}{t}(\mathbf{1}_{s_t} - z_{t-1}).$$

Taking the expectation over s_t—whose distribution is q_t—leads to (2.1).

3. Regret-based strategies

3.1. Regrets and the Hannan set

Given a joint distribution $z \in \Delta(S)$, the *regrets* of player i are defined by[6]

$$D_k^i(z) := u^i\big(k, z^{-i}\big) - u^i(z), \quad \text{for each } k \in S^i;$$

put $D^i(z) := (D_k^i(z))_{k \in S^i}$ for the *vector* of regrets.

It is useful to introduce the concept of the *Hannan set* H (of a given game Γ) as the set of all $z \in \Delta(S)$ satisfying

$$u^i(z) \geqslant \max_{k \in S^i} u^i\big(k, z^{-i}\big) \quad \text{for all } i \in N$$

(recall that z^{-i} denotes the marginal of z on S^{-i}); i.e., $z \in H$ if all regrets of all players are non-positive: $D^i(z) \leqslant 0$ for all $i \in N$. Thus, a joint distribution of actions lies in the Hannan set if the payoff of each player is no less than his best-reply payoff against the joint

[2] We thus view $\prod_{i \in N} \Delta(S^i)$ as the subset of independent distributions in $\Delta(S)$.

[3] If in fact the players play independently then $q(t) \in \prod_{i \in N} \Delta(S^i) \subset \Delta(S)$.

[4] Note that if $z(t)$ is on the boundary of $\Delta(S)$, i.e., if $(z(s))(t) = 0$ for some $s \in S$, then (2.1) implies $(\dot{z}(s))(t) \geqslant 0$, and thus $z(t)$ can never leave $\Delta(S)$.

[5] We write $\mathbf{1}_s$ for the unit vector in $\Delta(S)$ corresponding to the pure $s \in S$.

[6] It is convenient to extend multilinearly the payoff functions u^i from S to $\Delta(S)$, in fact to all $\mathbb{R}^S$; i.e., $u(z) := \sum_{s \in S} z(s)u(s)$ for all $z \in \mathbb{R}^S$. We slightly abuse notation and write expressions of the form (k, z^{-i}) or $k \times z^{-i}$ instead of $e_k^i \times z^{-i}$, where $k \in S^i$ and $e_k^i \in \Delta(S^i)$ is the k-unit vector.

distribution of actions of the other players (in the context of a repeated game, this is the Hannan (1957) condition).

We note that:

- The Hannan set H is a convex set (in fact a convex polytope).
- The Hannan set H contains all correlated equilibria,[7] and thus *a fortiori* all Nash equilibria.
- If z is independent over the players, then z is in the Hannan set H if and only if z is a Nash equilibrium.

3.2. Potential functions

General regret-based strategies make use of potential functions, introduced in Hart and Mas-Colell (2001a). A *potential function* on $\mathbb{R}^m$ is a function $P : \mathbb{R}^m \to \mathbb{R}$ satisfying:

(P1) P is a C^1 function; $P(x) > 0$ for all $x \notin \mathbb{R}^m_-$, and $P(x) = 0$ for all $x \in \mathbb{R}^m_-$;
(P2) $\nabla P(x) \geqslant 0$ and $\nabla P(x) \cdot x > 0$ for all $x \notin \mathbb{R}^m_-$;
(P3) $P(x) = P([x]_+)$ for[8] all x; and
(P4) there exist $0 < \rho_1 \leqslant \rho_2 < \infty$ such that $\rho_1 P(x) \leqslant \nabla P(x) \cdot x \leqslant \rho_2 P(x)$ for all $x \notin \mathbb{R}^m_-$.

Note that (P1), (P2), and (P3) correspond to (R1), (R2), and (R3) of Hart and Mas-Colell (2001a) for[9,10] $C = \mathbb{R}^m_-$. Condition (P4) is technical.[11]

The potential function P may be viewed as a generalized distance to $\mathbb{R}^m_-$; for example, take $P(x) = \min\{(\|x - y\|_p)^p\colon\ y \in \mathbb{R}^m_-\} = (\|[x]_+\|_p)^p$ where $\|\cdot\|_p$ is the l^p-norm on $\mathbb{R}^m$ and $1 < p < \infty$.

From now on we will always assume (P1)–(P4). By (P2), the gradient of P at $x \notin \mathbb{R}^m_-$ is a non-negative and non-zero vector; we introduce the notation[12]

$$\widehat{\nabla} P(x) := \frac{1}{\|\nabla P(x)\|} \nabla P(x) \in \Delta(m) \tag{3.1}$$

[7] Consider the setup where players get 'recommendations' before the play of the game. Correlated equilibria are those outcomes where no player can unilaterally gain by deviating from some recommendation. If only *constant* deviations (i.e., playing a fixed action regardless of the recommendation) are allowed, this yields the Hannan set. Note that if every player has two strategies, then the Hannan set coincides with the set of correlated equilibria. See Section 5.

[8] We write $[\xi]_+$ for the positive part of the real ξ, i.e., $[\xi]_+ = \max\{\xi, 0\}$; for a vector $x = (x_1, \ldots, x_m)$, we write $[x]_+$ for $([x_1]_+, \ldots, [x_m]_+)$.

[9] The second part of (P1) is without loss of generality—see Lemma 2.3(c1) and the construction of P_1 in the Proof of Theorem 2.1 of Hart and Mas-Colell (2001a).

[10] The 'better play' condition (R3) is 'If $x_k < 0$ then $\nabla_k P(x) = 0$,' which indeed implies that $P(x) = P([x]_+)$.

[11] $\nabla P(x) \cdot x / P(x) = \mathrm{d}P(\tau x)/\mathrm{d}\tau$ evaluated at $\tau = 1$; therefore it may be interpreted as the 'local returns to scale of P at x.' Condition (P4) thus says that the local returns to scale are uniformly bounded from above and from below (away from 0). If P is homogeneous of degree α then one can take $\rho_1 = \rho_2 = \alpha$.

[12] It will be convenient to use throughout the l^1-norm $\|x\| = \sum_k |x_k|$. The partial derivative $\partial P(x)/\partial x_k$ of $P(x)$ with respect to x_k is denoted $\nabla_k P(x)$ (it is the k-coordinate of the gradient vector $\nabla P(x)$). Finally, we write $\Delta(m)$ for the unit simplex of $\mathbb{R}^m$.

for the *normalized gradient* of P at x; thus $\widehat{\nabla}_k P(x) := \nabla_k P(x)/(\sum_{\ell \in S^i} \nabla_\ell P(x))$ for each $k = 1, \ldots, m$.

3.3. Regret-based strategies

'Regret-matching' is a repeated game strategy where the probabilities of play are proportional to the positive part of the regrets (i.e., to $[D^i(z)]_+$). This is a special case of what we will call regret-based strategies.

We say that player i uses a *regret-based strategy* if there exists a potential function $P^i : \mathbb{R}^{S^i} \to \mathbb{R}$ (satisfying (P1)–(P4)) such that at each time t where some regret of player i is positive, the mixed play $q^i(t) \in \Delta(S^i)$ of i is proportional to the gradient of the potential evaluated at the current regret vector; that is,

$$q^i(t) = \widehat{\nabla} P^i\big(D^i\big(z(t)\big)\big) \quad \text{when } D^i\big(z(t)\big) \notin \mathbb{R}_-^{S^i}. \tag{3.2}$$

Note that there are no conditions when the regret vector is non-positive. Such a strategy is called a P^i*-strategy* for short.

Condition (3.2) is the counterpart of the discrete-time P^i-strategy of Hart and Mas-Colell (2001a):

$$q_k^i(T+1) \equiv \Pr\big[s_{T+1}^i = k \mid h_T\big] = \widehat{\nabla} P_k^i\big(D^i(z_T)\big) \quad \text{when } D^i(z_T) \notin \mathbb{R}_-^{S^i}.$$

Remark. The class of regret-based strategies of a player i is invariant to transformations of i's utility function which preserve i's mixed-action best-reply correspondence (i.e., replacing u^i with $\tilde{u}^i$ given by $\tilde{u}^i(s) := \alpha u^i(s) + v(s^{-i})$ for some $\alpha > 0$; indeed, $v(\cdot)$ does not affect the regrets, and α changes the scale, which requires a corresponding change in P^i).

The main property of regret-based strategies (see Hart and Mas-Colell, 2001a, Theorem 3.3, for the discrete-time analog) is:

Theorem 3.1. *Let $z(t)$ be a solution of* (2.1) *and* (3.2)*. Then $\overline{\lim}_{t\to\infty} D_k^i(z(t)) \leqslant 0$ for every $k \in S^i$.*

Remark. This result holds for *any* strategies of the other players q^{-i}; in fact, one may allow correlation between the players in $N\backslash\{i\}$ (but, of course, q^{-i} must be independent of q^i—thus $q(t) = q^i(t) \times q^{-i}(t)$).

Proof. For simplicity rescale the time t so that (2.1) becomes[13] $\dot{z} = q - z$. Assume $D^i(z) \notin \mathbb{R}_-^{S^i}$, so $P^i(D^i(z)) > 0$. We have (recall Footnote 6)

$$\begin{aligned}
\dot{D}_k^i(z) &= u^i\big(k \times \dot{z}^{-i} - \dot{z}\big) = u^i\big(k \times \big(q^{-i} - z^{-i}\big) - q^i \times q^{-i} + z\big)\\
&= u^i\big(k \times q^{-i} - q^i \times q^{-i} - k \times z^{-i} + z\big)\\
&= u^i\big(k, q^{-i}\big) - u^i\big(q^i, q^{-i}\big) - D_k^i(z).
\end{aligned}$$

[13] Take $\tilde{t} = \exp(t)$.

Multiplying by q_k^i and summing over $k \in S^i$ yields

$$q^i \cdot \dot{D}^i(z) = -q^i \cdot D^i(z). \tag{3.3}$$

Define $\pi^i(z) := P^i(D^i(z))$; then (recall (3.2))

$$\begin{aligned}\dot{\pi}^i(z) &= \nabla P^i\big(D^i(z)\big) \cdot \dot{D}^i(z) = \big\|\nabla P^i\big(D^i(z)\big)\big\| q^i \cdot \dot{D}^i(z)\\ &= -\big\|\nabla P^i\big(D^i(z)\big)\big\| q^i \cdot D^i(z) = -\nabla P^i\big(D^i(z)\big) \cdot D^i(z).\end{aligned} \tag{3.4}$$

Using condition (P2) implies that $\dot{\pi}^i < 0$ when $D^i(z) \notin \mathbb{R}_-^{S^i}$—thus π^i is a strict Lyapunov function for the dynamical system. It follows that[14] $\pi^i(z) \to 0$. □

Corollary 3.2. *If all players play regret-based strategies, then $z(t)$ converges as $t \to \infty$ to the Hannan set H.*

One should note that here (as in all the other results of this paper), the convergence is to the *set* H, and not to a specific point in that set. That is, the distance between $z(t)$ and the set H converges to 0; or, equivalently, the limit of any convergent subsequence lies in the set.

We end this section with a technical result: Once there is some positive regret, then a regret-based strategy will maintain this forever (of course, the regrets go to zero by Theorem 3.1).

Lemma 3.3. *If $D^i(z(t_0)) \notin \mathbb{R}_-^{S^i}$ then $D^i(z(t)) \notin \mathbb{R}_-^{S^i}$ for all $t \geqslant t_0$.*

Proof. Let $\pi^i := P^i(D^i(z))$. Then $\pi^i(t_0) > 0$, and (3.4) together with (P3) implies that $\dot{\pi}^i \geqslant -\rho_2 \pi^i$ and thus $\pi^i(t) \geqslant \mathrm{e}^{-\rho_2(t-t_0)} \pi^i(t_0) > 0$ for all $t > t_0$. □

4. Nash equilibria

In this section we consider two-person games, and show that in some special classes of games regret-based strategies by both players do in fact lead to the set of Nash equilibria (not just to the Hannan set, which is in general a strictly larger set).

If z belongs to the Hannan set H, then $u^i(z) \geqslant u^i(k^i, z^j)$ for all $k^i \in S^i$ and $i \neq j$. Averaging according to z^i yields

$$u^i(z) \geqslant u^i\big(z^1, z^2\big) \quad \text{for } i = 1, 2. \tag{4.1}$$

Lemma 4.1. *In a two-person game, if z belongs to the Hannan set and the payoff of z is the same as the payoff of the product of its marginals, i.e., if*

$$u^i(z) = u^i\big(z^1, z^2\big) \quad \textit{for } i = 1, 2, \tag{4.2}$$

then (z^1, z^2) is a Nash equilibrium.

[14] Note that only (P1) and (P2) were used in this proof.

Proof. If $z \in H$ then $u^i(k^i, z^j) \leqslant u^i(z) = u^i(z^1, z^2)$ for all $k^i \in S^i$. $\square$

4.1. Two-person zero-sum games

Consider a two-person zero-sum game Γ, i.e., $u^1 = u$ and $u^2 = -u$. Let v denote the minimax *value* of Γ. A pair of (mixed) strategies (y^1, y^2) is a Nash equilibrium if and only if y^i is an *optimal* strategy of player i (i.e., if it guarantees the value v).

Theorem 4.2. *Let Γ be a two-person zero-sum game. If both players play regret-based strategies, then $(z^1(t), z^2(t))$ converges to the set of Nash equilibria of Γ, and $u(z(t))$ and $u(z^1(t), z^2(t))$ both converge as $t \to \infty$ to the minimax value v of Γ.*

Proof. The inequalities (4.1) for both players imply the equalities (4.2), and the result follows from Theorem 3.1 and Lemma 4.1. $\square$

See Corollary 4.5 in Hart and Mas-Colell (2001a) for the discrete-time analog.

4.2. Two-person potential games

Consider a two-person *potential game* Γ: Without loss of generality the two players have identical payoff functions[15] $u^1 = u^2 = u : S \to \mathbb{R}$.

We will show first that if initially[16] each player has some positive regret, then both players using regret-based strategies leads to the set of Nash equilibria. Regret-based strategies allow a player to behave arbitrarily when all his regrets are non-positive—in particular, inside the Hannan set (which is larger than the set of Nash equilibria). In order to extend our result and always guarantee convergence to the set of Nash equilibria, the strategies need to be appropriately defined in the case of non-positive regrets; we do so at the end of this subsection.

Before proceeding we need a technical lemma.

Lemma 4.3. *Let P be a potential function* (*satisfying* (P1)–(P4)). *Then for every $K > 0$ there exists a constant $c > 0$ such that*

$$\max_k x_k \leqslant c\big(P(x)\big)^{1/\rho_2} \quad \textit{for all } x \in [-K, K]^m.$$

Proof. Since replacing P with P^{1/ρ_2} does not affect (P1)–(P4), we can assume without loss of generality that $\rho_2 = 1$ in (P4). Take a non-negative $x \in [0, K]^m$, and let $f(\tau) := P(\tau x)$ for $\tau \geqslant 0$. Then $f'(\tau) = \nabla P(\tau x) \cdot x \leqslant P(\tau x)/\tau = f(\tau)/\tau$ for all $\tau > 0$; hence $(f(\tau)/\tau)' \leqslant 0$, which implies that $f(\tau)/\tau \geqslant f(1)$ for all $\tau \leqslant 1$. Thus $P(\tau x) \geqslant \tau P(x)$ for all $x \geqslant 0$ and all $0 \leqslant \tau \leqslant 1$. Let $a := \min\{P(x)\colon\ x \geqslant 0,\ \|x\| = K\}$, then $a > 0$ since the minimum is attained. Hence

$$P(x) = P(x_1, \ldots, x_m) \geqslant P(x_1, 0, \ldots, 0) \geqslant \frac{x_1}{K} P(K, 0, \ldots, 0) \geqslant \frac{x_1}{K} a$$

[15] Recall the remark preceding Theorem 3.1.

[16] I.e., at $t = 1$—or, in fact, at any $t = t_0$.

(the first inequality since $\nabla P \geqslant 0$). Altogether we get $x_1 \leqslant cP(x)$ where $c = K/a$; the same applies to the other coordinates. For $x \in [-K, K]^m$ which is not non-negative, use (P3):

$$\max_k x_k \leqslant \max_k [x_k]_+ \leqslant cP\big([x]_+\big) = cP(x).$$

This completes the proof. □

By replacing P with cP^{1/ρ_2} for an appropriate $c > 0$—which does not affect the normalized gradient—we will assume from now on without loss of generality that the potential P^i for each player i is chosen so that

$$\max_{k \in S^i} x_k \leqslant P^i(x) \quad \text{for all } x \in [-2M, 2M]^{S^i}. \tag{4.3}$$

We deal first with the case where initially, at $t = 1$, both players have some positive regret.

Theorem 4.4. *Let Γ be a two-person potential game. Assume that initially both players have some positive regret, i.e., $D^i(z(1)) \notin \mathbb{R}_-^{S^i}$ for $i = 1, 2$. If both players use regret-based strategies, then the pair of marginal distributions $(z^1(t), z^2(t)) \in \Delta(S^1) \times \Delta(S^2)$ converges as $t \to \infty$ to the set of Nash equilibria of the game. Moreover, there exists a number $\bar{v}$ such that $(z^1(t), z^2(t))$ converges to the set of Nash equilibria with payoff $\bar{v}$* (*to both players*), *and the average payoff $u(z(t))$ also converges to $\bar{v}$.*

Proof. We again rescale t so that $\dot{z} = q - z$. Lemma 3.3 implies that $\pi^i(t) := P^i(D^i(z^i(t))) > 0$ for all t. We have

$$\begin{aligned} \dot{u}\big(z^1, z^2\big) &= \dot{u}\big(z^1 \times z^2\big) = u\big(\dot{z}^1 \times z^2 + z^1 \times \dot{z}^2\big) \\ &= u\big(\big(q^1 - z^1\big) \times z^2 + z^1 \times \big(q^2 - z^2\big)\big) \\ &= u\big(q^1, z^2\big) + u\big(z^1, q^2\big) - 2u\big(z^1, z^2\big). \end{aligned}$$

Now

$$u\big(q^1, z^2\big) = \sum_{k \in S^1} q_k^1 u\big(k, z^2\big) = u(z) + \sum_{k \in S^1} q_k^1 D_k^1(z) = u(z) + q^1 \cdot D^1 > u(z)$$

(by (P2) since q^1 is proportional to $\nabla P^1(D^1)$). Thus

$$\dot{u}\big(z^1, z^2\big) > 2u(z) - 2u\big(z^1, z^2\big). \tag{4.4}$$

Next, (4.3) implies

$$u\big(k, z^2\big) - u(z) = D_k^1(z) \leqslant P^1\big(D^1(z)\big) = \pi^1$$

for all $k \in S^1$, and therefore

$$u\big(z^1, z^2\big) - u(z) \leqslant \pi^1. \tag{4.5}$$

Similarly for player 2, and thus from (4.4) we get

$$\dot{u}\big(z^1, z^2\big) > -\pi^1 - \pi^2. \tag{4.6}$$

Now

$$\dot{\pi}^i = -\nabla P^i\big(D^i(z)\big)\cdot D^i(z) \leqslant -\rho P^i\big(D^i(z)\big) = -\rho\pi^i \tag{4.7}$$

(we have used (3.4) and (P3), with ρ the minimum of ρ_1^i of (P4) for $i = 1, 2$).

Define $v := u(z^1, z^2) - \pi^1/\rho - \pi^2/\rho$; from (4.6) we get

$$\dot{v} = \dot{u}\big(z^1, z^2\big) - \dot{\pi}^1/\rho - \dot{\pi}^2/\rho \geqslant \dot{u}\big(z^1, z^2\big) + \pi^1 + \pi^2 > 0. \tag{4.8}$$

Therefore v increases; since it is bounded, it converges; let $\bar{v}$ be its limit. Theorem 3.1 implies that $\pi^i \to 0$, so $u(z^1, z^2) \to \bar{v}$.

By Lemma 4.1, it remains to show that $u(z) - u(z^1, z^2) \to 0$. We use the following

Lemma 4.5. *Let $f:\mathbb{R}_+\to\mathbb{R}_+$ be a non-negative, uniformly Lipschitz function such that $\int_0^\infty f(t)\,\mathrm{d}t < \infty$. Then $f(t) \to 0$ as $t \to \infty$.*

Proof. Let L be such that $|f(t) - f(T)| \leqslant L|t - T|$ for all t, T. If $f(T) \geqslant 2\varepsilon > 0$, then $f(t) \geqslant \varepsilon$ for all $T \leqslant t \leqslant T + \varepsilon/L$, so $\int_T^{T+\varepsilon/L} f(t)\,\mathrm{d}t \geqslant \varepsilon^2/L$. Since the integral is bounded, it follows that there can be at most finitely many such occurrences, so $f(T) < 2\varepsilon$ for all T large enough. □

To get back to the proof of Theorem 4.4: define $f := 2u(z) - 2u(z^1, z^2) + \pi^1 + \pi^2$; then f is non-negative (by (4.5)) and uniformly Lipschitz, and $\int_0^\infty f(t)\,\mathrm{d}t$ is finite (it is bounded by $\bar{v}$, since $f < \dot{u}(z^1, z^2) + \pi^1 + \pi^2 \leqslant \dot{v}$ by (4.4) and (4.8)). Lemma 4.5 implies that $f \to 0$; thus $u(z) - u(z^1, z^2) \to 0$ (since $\pi^i \to 0$). □

We handle now the case where at the initial condition $z(1)$ all the regrets of a player i are non-positive. We define the strategy of i as follows: i plays an arbitrary *fixed* mixed strategy $\bar{y}^i \in \Delta(S^i)$, up to such time T^i when some regret is at least $1/T^i$ (i.e., T^i is the first $t > 1$ such that[17] $\max_{k\in S^i} D_k^i(z(t)) \geqslant 1/t$); of course, if this never happens (i.e., if $T^i = \infty$), then i always plays $\bar{y}^i$. After time T^i player i plays P^i-regret-matching (recall Lemma 3.3). That is,

$$q^i(t) := \begin{cases} \bar{y}^i, & \text{for } t \leqslant T^i, \\ \widehat{\nabla} P^i(D^i(z(t))) & \text{for } t > T^i. \end{cases} \tag{4.9}$$

Corollary 4.6. *The result of Theorem* 4.4 *holds for any initial* $z(1)$ *when the strategies are given by* (4.9).

Proof. If there is some time T after which both players play $q^i = \widehat{\nabla} P^i(D^i(z))$, then we apply Theorem 4.4 starting at T. Otherwise, for a player i that plays $\bar{y}^i$ forever, we have $\max_{k\in S^i} D_k^i(z(t)) < 1/t$ for all t, so $D^i(z(t)) \to \mathbb{R}_-^{S^i}$. Moreover $z^i(t)$ converges to the constant $\bar{y}^i$ and so $z(t)$ becomes independent in the limit (i.e., $z(t) - z^i(t) \times z^{-i}(t) \to 0$);

[17] We use $1/t$ rather than 0 in order to avoid difficulties at the boundary of $\mathbb{R}_-^{S^i}$; any positive function of t converging to 0 as $t \to \infty$ will do.

the convergence to the set of Nash equilibria follows from Lemma 4.1. Finally, the payoff $\bar{v}$ is just the best-reply payoff against $\bar{y}^i$. □

The analog of this result for discrete-time—which is a new result—is stated and proved in Appendix A.

4.3. Other classes of games

Smooth fictitious play—which may be viewed as (approximately) a limiting case of regret-based strategies—has been shown to converge to the set of (approximate) Nash equilibria for additional classes of two-person games, namely, games with a unique interior ESS, and supermodular games (see Hofbauer and Sandholm, 2002). It turns out that general regret-based strategies converge to Nash equilibria for the first class (Hofbauer, personal communication (2002)); we do not know about the second class.

5. Correlated equilibria

Given a joint distribution $z \in \Delta(S)$, the regret of player i for action k may be rewritten as follows:

$$\begin{aligned} D^i_k(z) &= \sum_{s \in S} \left[u^i\left(k, s^{-i}\right) - u^i(s)\right] z(s) \\ &= \sum_{j \in S^i} \sum_{s^{-i} \in S^{-i}} \left[u^i\left(k, s^{-i}\right) - u^i\left(j, s^{-i}\right)\right] z\left(j, s^{-i}\right). \end{aligned}$$

We now define the *conditional regret of player i from action j to action k* (for $j, k \in S^i$ with $j \neq k$) as follows:

$$C^i_{jk}(z) := \sum_{s^{-i} \in S^{-i}} \left[u^i\left(k, s^{-i}\right) - u^i\left(j, s^{-i}\right)\right] z\left(j, s^{-i}\right). \tag{5.1}$$

This is the change in the payoff of i if action j had always been replaced by action k. Denote $L := \{(j,k) \in S^i \times S^i\colon\ j \neq k\}$ and let $C^i(z) := (C^i_{jk}(z))_{(j,k) \in L}$ be the vector of conditional regrets. A distribution $z \in \Delta(S)$ is a *correlated equilibrium* if and only if $C^i(z) \leqslant 0$ for all $i \in N$ (see Hart and Mas-Colell, 2000).[18]

Conditional regret-based strategies for a player i will define the action of i by the way it changes with time—i.e., by a differential equation. This requires us to add $q^i(t) \in \Delta(S^i)$ as a state variable—in addition to $z(t) \in \Delta(S)$, which changes according to (2.1). Specifically, we say that player i plays a *conditional regret-based strategy* if there exists a potential function $P^i : \mathbb{R}^L \to \mathbb{R}$ (satisfying (P1)–(P4)), such that, when $C^i(z(t)) \notin \mathbb{R}^L_-$,

$$\dot{q}^i_j(t) = \sum_{k \neq j} \nabla_{(k,j)} P^i\left(C^i\left(z(t)\right)\right) q^i_k(t) - \sum_{k \neq j} \nabla_{(j,k)} P^i\left(C^i\left(z(t)\right)\right) q^i_j(t) \tag{5.2}$$

[18] Note that $C^i_{jk}(z) \leqslant 0$ for all $j \neq k$ implies $D^i_k(z) = \sum_{j \neq k} C^i_{jk}(z) \leqslant 0$; this shows that the Hannan set contains the set of correlated equilibria (recall Section 3.1 and Footnote 7).

for all $j \in S^i$, where $\nabla_{(k,j)}$ denotes the derivative with respect to the (k, j)-coordinate;[19] again, there are no conditions when all conditional regrets are non-positive, i.e., when $C^i(z(t)) \in \mathbb{R}^L_-$.

To see where (5.2) comes from, recall the discrete-time strategy of Hart and Mas-Colell (2000, (2.2)):

$$q^i_j(t+1) = \mathbf{1}_{s^i_t=j}\left[1 - \frac{1}{\mu}\sum_{k\neq j} R^i_{(j,k)}(t)\right] + \sum_{k\neq j} \mathbf{1}_{s^i_t=k}\frac{1}{\mu}R^i_{(k,j)}(t),$$

which, when taking expectations, yields

$$\begin{aligned} q^i_j(t+1) &= q^i_j(t)\left[1 - \frac{1}{\mu}\sum_{k\neq j} R^i_{(j,k)}(t)\right] + \sum_{k\neq j} q^i_k(t)\frac{1}{\mu}R^i_{(k,j)}(t) \\ &= q^i_j(t) + \frac{1}{\mu}\sum_{k\neq j}\left[R^i_{(k,j)}(t)q^i_k(t) - R^i_{(j,k)}(t)q^i_j(t)\right]. \end{aligned}$$

Replacing the positive part of the regrets $R^i_{(k,j)} = [C^i_{(j,k)}]_+$ with their generalizations $\nabla_{(k,j)}P^i(C^i)$ leads to (5.2) (see Hart and Mas-Colell, 2001a, Section 5.1).

Remarks.

(1) The ‘speeds of adjustment’ of q and z (a constant for q, and $1/t$ for z) are different.

(2) We have $\sum_j \dot{q}^i_j = 0$ and $\dot{q}^i_j \geqslant 0$ when $q^i_j = 0$; therefore q^i never leaves the simplex $\Delta(S^i)$ if we start there (i.e., if $q^i(1) \in \Delta(S^i)$).

Theorem 5.1. *If player i plays a conditional regret-based strategy, then*

$$\overline{\lim}_{t\to\infty} \max_{j,k} C^i_{jk}\big(z(t)\big) \leqslant 0$$

for any play $q^{-i}(t)$ of the other players.

Corollary 5.2. *If all players use conditional regret-based strategies, then $z(t)$ converges as $t \to \infty$ to the set of correlated equilibria of the game Γ.*

Remark. Unlike the discrete-time case (see the discussion in Hart and Mas-Colell, 2000, Section 4(d)), the result for continuous time applies to each player separately; that is, no assumption is needed on q^{-i} in Theorem 5.1. The reason is that, in the ‘limit’—as the time periods become infinitesimal—the condition of Cahn (2000) is essentially satisfied by any continuous solution.[20] Thus continuous-time conditional regret-based strategies are

[19] (5.2) may be viewed as the differential equation for the expected probability of a continuous-time Markov process.

[20] The Cahn condition is that the effect of the choice of player i at time t on the choice of another player j at some future time goes to zero as t goes to infinity. More precisely, if the histories h_{t+w-1} and h'_{t+w-1} differ *only* in their s^i_t-coordinate, then for all $j \neq i$ we have $|\Pr[s^j_{t+w} = s^j \mid h_{t+w-1}] - \Pr[s^j_{t+w} = s^j \mid h'_{t+w-1}]| \leqslant f(w)/g(t)$ for some functions f and g such that $g(t) \to 0$ as $t \to \infty$.

'universally conditionally consistent' or 'universally calibrated' (cf. Fudenberg and Levine, 1998, 1999).

Proof of Theorem 5.1. Assume without loss of generality that in (P4) we have $\rho_1 = 1$ (replace P^i with $(P^i)^{1/\rho_1}$). Throughout this proof, j and k will always be elements of S^i; we have:

$$\begin{aligned}\dot{C}^i_{jk}(z) &= \sum_{s^{-i}\in S^{-i}} \bigl[u^i\bigl(k,s^{-i}\bigr) - u^i\bigl(j,s^{-i}\bigr)\bigr]\dot{z}\bigl(j,s^{-i}\bigr)\\ &= \frac{1}{t}\sum_{s^{-i}\in S^{-i}} \bigl[u^i\bigl(k,s^{-i}\bigr) - u^i\bigl(j,s^{-i}\bigr)\bigr]\bigl(-z\bigl(j,s^{-i}\bigr) + q^i_j q^{-i}_{s^{-i}}\bigr)\\ &= \frac{1}{t}\Bigl(-C^i_{jk}(z) + \sum_{s^{-i}\in S^{-i}} \bigl[u^i\bigl(k,s^{-i}\bigr) - u^i\bigl(j,s^{-i}\bigr)\bigr]q^i_j q^{-i}_{s^{-i}}\Bigr). \qquad (5.3)\end{aligned}$$

Denote $\pi(t) := P^i(C^i(z(t)))$ and $G(t) = \nabla P^i(C^i(z(t)))$. Then

$$\begin{aligned}\dot{\pi} &= G\cdot \dot{C}^i(z)\\ &\leqslant -\frac{1}{t}\pi + \frac{1}{t}\sum_{s^{-i}\in S^{-i}} q^{-i}_{s^{-i}}\Bigl\{\sum_{j,k} G_{jk}\bigl[u^i\bigl(k,s^{-i}\bigr) - u^i\bigl(j,s^{-i}\bigr)\bigr]q^i_j\Bigr\}, \qquad (5.4)\end{aligned}$$

where we have used (P4) (recall that $\rho_1 = 1$). Denote by E the right-hand sum over s^{-i}, and by $E(s^{-i})$ the expression in the curly brackets $\{\ldots\}$ (thus E is a weighted average of the $E(s^{-i})$). Rearranging terms yields

$$E\bigl(s^{-i}\bigr) = \sum_j u\bigl(j,s^{-i}\bigr)\Bigl[\sum_k G_{kj}q^i_k - \sum_k G_{jk}q^i_j\Bigr] = \sum_j u\bigl(j,s^{-i}\bigr)\dot{q}^i_j.$$

We now claim that $\dot{q}^i_j \to 0$ as $t\to\infty$ for all $j\in S^i$.

Indeed, let $m := |S^i|$, then $|C^i_{jk}| \leqslant 2M$ and so $\|C^i\| \leqslant 2Mm(m-1) =: M_1$; also $|\dot{C}^i_{jk}| \leqslant 2M|S^{-i}|/t$ (since $|\dot{z}(s)| \leqslant 1/t$ for all s by (2.1)) and thus

$$\bigl\|\dot{C}^i\bigr\| \leqslant 2M\bigl|S^{-i}\bigr|m(m-1)/t =: M_2/t.$$

Let K be a Lipschitz bound for $\nabla P^i(x)$ over $\|x\| \leqslant M_1$; then for all $t_2 \geqslant t_1 \geqslant 1$ we have

$$\begin{aligned}\bigl\|G(t_2) - G(t_1)\bigr\| &\leqslant K\bigl\|C^i\bigl(z(t_2)\bigr) - C^i\bigl(z(t_1)\bigr)\bigr\| \leqslant K\bigl\|\dot{C}^i\bigl(z(\tau)\bigr)\bigr\|(t_2-t_1)\\ &\leqslant KM_2\frac{t_2-t_1}{t_1} \qquad (5.5)\end{aligned}$$

($\tau\in[t_1,t_2]$ is some intermediate point).

Let $M_3 := \max_{\|x\|\leqslant M_1}\|\nabla P^i(x)\|$, and define

$$A_{jk}(t) := \frac{1}{M_3}G_{jk}(t), \quad \text{for } j\neq k, \quad \text{and}$$

$$A_{jj}(t) := 1 - \frac{1}{M_3}\sum_{k\neq j} G_{kj}(t).$$

Then $A(t)$ is a stochastic matrix,[21] and (5.2) can be rewritten as[22]

$$\dot{q}^i(t) = M_3 q^i(t)\big(A(t) - I\big). \tag{5.6}$$

Finally, (5.5) yields[23]

$$\big\|A(t_2) - A(t_1)\big\| \leqslant \frac{m}{M_3}\big\|G(t_2) - G(t_1)\big\| \leqslant \frac{mKM_2}{M_3}\frac{t_2 - t_1}{t_1}$$

for all $t_2 \geqslant t_1 \geqslant 1$.

Applying Proposition B.1 (see Appendix B; the constant M_3 in (5.6) does not matter—replace t by $M_3 t$) implies that indeed $\dot{q}^i \to 0$ as $t \to \infty$.

Therefore $E(s^{-i}) \to 0$ and so (recall (5.4)) $t\dot{\pi}(t) + \pi(t) \leqslant E(t) \to 0$ as $t \to \infty$, from which it follows that $\pi(t) \to 0$ (indeed, for each $\varepsilon > 0$ let $t_0 \equiv t_0(\varepsilon)$ be such that $|E(t)| \leqslant \varepsilon$ for all $t \geqslant t_0$; then $\mathrm{d}(t\pi(t))/\mathrm{d}t \leqslant \varepsilon$ for all $t \geqslant t_0$, which yields $t\pi(t) \leqslant t_0\pi(t_0) + \varepsilon(t - t_0)$ and thus $\overline{\lim}_{t\to\infty}\pi(t) \leqslant \varepsilon$). $\square$

6. Remarks

(a) It is worthwhile to emphasize, once again, that the appropriate state space for our analysis is not the product of the mixed action spaces of the players $\prod_i \Delta(S^i)$, but the space of joint distributions on the product of their pure action sets $\Delta(\prod_i S^i)$. This is so because, as we pointed out in Hart and Mas-Colell (2001a, Section 4), with the exception of the limiting case constituted by fictitious play, the dynamics of regret-matching depend on $u(z)$, the time-average of the realized payoffs, and therefore on the joint distribution z. It is interesting to contrast this with, for example, Hofbauer (2000) and Sandholm (2002), where, in an evolutionary context, dynamics similar to regret-matching are considered but where, nonetheless, the context dictates that the appropriate state space is the product of the mixed action spaces. This family of evolutionary dynamics is named by Hofbauer (2000) 'Brown–von Neumann–Nash dynamics.'

(b) The fact that the state space variable is the time-average distribution of play $z(t)$ does not impose on players informational requirements additional to those familiar from, say, fictitious play. It only asks that players record also their own play at each period (i.e., i keeps track of the frequency of each s, and not only of s^{-i}).

(c) One could ask to what extent the discrete-time analog of the results in this paper can be obtained by appealing to stochastic approximation techniques (see Benaïm, 1999, or Benaïm and Weibull, 2003). We have not investigated this matter in detail. However, it seems to us that for the results of Section 3 and Appendix C it should be a relatively simple matter, but for those of Sections 4 (Nash equilibria) and 5 (correlated equilibria) there may be a real challenge.

21 I.e., its elements are nonnegative and the sum of each row is 1.

22 Vectors (like q) are viewed as row vectors; I denotes the identity matrix.

23 The norm $\|A\|$ of a matrix A is taken to be $\max\{\|xA\|\colon \|x\| = 1\}$, so that always $\|xA\| \leqslant \|x\|\,\|A\|$. Note that if $A = (A_{jk})$ is an $m \times m$ matrix then $\max_{j,k}|A_{jk}| \leqslant \|A\| \leqslant m\max_{j,k}|A_{jk}|$; if moreover A is a stochastic matrix, then $\|A\| = 1$.

Acknowledgments

Research is partially supported by grants of the Israel Academy of Sciences and Humanities, the Spanish Ministry of Education, the Generalitat de Catalunya, and the EU-TMR Research Network. We thank Drew Fudenberg, Josef Hofbauer, Gil Kalai, David Levine, Abraham Neyman, Yosef Rinott, William Sandholm, and Benjamin Weiss for their comments and suggestions.

Appendix A. Discrete-time dynamics for potential games

In this appendix we deal with discrete-time dynamics for two-person potential games (see Section 4.2). We assume that the potential function P^i of each player satisfies (P1)–(P4) and, in addition,

(P5) P is a C^2 function.

A *discrete-time regret-based strategy* of player i is defined as follows: If $D^i(z_{t-1}) \notin \mathbb{R}_-^{S^i}$ (i.e., if there is some positive regret), then the play probabilities are proportional to the gradient of the potential $\nabla P^i(D^i(z_{t-1}))$. If $D^i(z_{t-1}) \in \mathbb{R}_-^{S^i}$ (i.e., if there is no positive regret),[24] then we assume that i uses the empirical distribution of his past choices z_{t-1}^i. One simple way to implement this is to choose at random a past period $r = 1, 2, \ldots, t-1$ (with equal probabilities of $1/(t-1)$ each) and play at time t the same action that was played at time r (i.e., $s_t^i = s_r^i$).[25] To summarize: At time t the action of player i is chosen according to the probability distribution $q_t^i \in \Delta(S^i)$ given by

$$q_t^i(k) = \Pr\left[s_t^i = k \mid h_{t-1}\right] := \begin{cases} \widehat{\nabla}_k P^i\left(D^i(z_{t-1})\right), & \text{if } D^i(z_{t-1}) \notin \mathbb{R}_-^{S^i}, \\ z_{t-1}^i(k), & \text{if } D^i(z_{t-1}) \in \mathbb{R}_-^{S^i}, \end{cases} \tag{A.1}$$

for each $k \in S^i$ (starting at $t = 1$ with an arbitrary $q_1^i \in \Delta(S^i)$).

Theorem A.1. *Let Γ be a two-person potential game. If both players use regret-based strategies* (A.1)*, then, with probability* 1, *the pair of empirical marginal distributions (z_t^1, z_t^2) converges as $t \to \infty$ to the set of Nash equilibria of the game, and the average realized payoff $u(z_t)$ (and $u(z_t^1, z_t^2)$) converges to the set of Nash equilibrium payoffs.*

Proof. Without loss of generality assume that (4.3) holds for both players (thus $\rho_2^i = 1$ in (P4)), and let $\rho > 0$ be the minimum of the ρ_1^i in (P4). Put $d_t^i(k) := D_k^i(z_t)$ for the k-regret and $d_t^i := D^i(z_t)$ for the vector of regrets, and $\pi_t^i := P^i(D^i(z_t)) = P^i(d_t^i)$. For clarity, we divide the proof into five steps.

Step 1. $\pi_t^i \to 0$ as $t \to \infty$ a.s., and there exists a constant M_1 such that

$$\mathrm{E}\left[\pi_t^i \mid h_{t-1}\right] \leqslant (1 - \rho/t)\pi_{t-1}^i + \frac{M_1}{t^2}. \tag{A.2}$$

[24] Unlike the continuous-time case (recall Lemma 3.3), here the regret vector may enter and exit the negative orthant infinitely often—which requires a more delicate analysis.

[25] In short: There is no change when there is no regret. Other definitions are possible in this case of ‘no regret’—for example, the result of Theorem A.1 can be shown to hold also if a player plays optimally against the empirical distribution of the other player (i.e., ‘fictitious play’) when all his regrets are non-positive.

Proof. [26] Consider player 1; for each $k \in S^1$ we have

$$\begin{aligned}\mathrm{E}\big[d_t^1(k) - d_{t-1}^1(k) \mid h_{t-1}\big] &= \frac{t-1}{t}u\big(k, z_{t-1}^2\big) + \frac{1}{t}u\big(k, q_t^2\big) - \frac{t-1}{t}u(z_{t-1}) - \frac{1}{t}u\big(q_t^1, q_t^2\big) \\ &\quad - u\big(k, z_{t-1}^2\big) + u(z_{t-1}) \\ &= \frac{1}{t}\big(u\big(k, q_t^2\big) - u\big(q_t^1, q_t^2\big)\big) - \frac{1}{t}d_{t-1}^1(k).\end{aligned}$$

The first term vanishes when averaging according to q_t^1, so

$$\mathrm{E}\big[q_t^1 \cdot \big(d_t^1 - d_{t-1}^1\big) \mid h_{t-1}\big] = -\frac{1}{t}q_t^1 \cdot d_{t-1}^1$$

(compare with (3.3)). If $d_{t-1}^1 \notin \mathbb{R}_-^{S^1}$ then q_t^1 is proportional to $\nabla P^1(d_{t-1}^1)$; hence

$$\mathrm{E}\big[\nabla P^1\big(d_{t-1}^1\big) \cdot \big(d_t^1 - d_{t-1}^1\big) \mid h_{t-1}\big] = -\frac{1}{t}\nabla P^1\big(d_{t-1}^1\big) \cdot d_{t-1}^1 \leqslant -\frac{\rho}{t}P^1\big(d_{t-1}^1\big)$$

by (P4). This also holds when $d_{t-1}^1 \in \mathbb{R}_-^{S^1}$ (since then both P^1 and ∇P^1 vanish). Therefore, by (P5), there exists some constant M_1 such that

$$\mathrm{E}\big[P^1\big(d_t^1\big) - P^1\big(d_{t-1}^1\big) \mid h_{t-1}\big] \leqslant -\frac{\rho}{t}P^1\big(d_{t-1}^1\big) + \frac{M_1}{t^2},$$

which is (A.2). Finally, $\pi_t^i \to 0$ follows from Theorem 3.3 in Hart and Mas-Colell (2001a) (or use (A.2) directly).

Step 2. Let[27] $\alpha_{t-1}^i := u(q_t^i, z_{t-1}^j) - u(z_{t-1}^1, z_{t-1}^2) + \pi_{t-1}^i$. Then $\alpha_{t-1}^i \geqslant 0$ and moreover:

$$\text{If} \quad \pi_{t-1}^i > 0 \quad \text{then} \quad \alpha_{t-1}^i > u(z_{t-1}) - u\big(z_{t-1}^1, z_{t-1}^2\big) + \pi_{t-1}^i \geqslant 0. \tag{A.3}$$

Proof. Take $i = 1$. We have

$$u\big(k, z_{t-1}^2\big) - u(z_{t-1}) = d_{t-1}^1(k) \leqslant P^1\big(d_{t-1}^1\big) = \pi_{t-1}^1 \tag{A.4}$$

for all $k \in S^1$ by (4.3). Averaging over k according to z_{t-1}^1 yields

$$u\big(z_{t-1}^1, z_{t-1}^2\big) - u(z_{t-1}) \leqslant \pi_{t-1}^1.$$

If $\pi_{t-1}^1 = 0$ then $q_t^1 = z_{t-1}^1$ and so $\alpha_{t-1}^1 = 0$. If $\pi_{t-1}^1 > 0$ then $q_t^1 \cdot d_{t-1}^1 = \widehat{\nabla} P^1(d_{t-1}^1) \cdot d_{t-1}^1 > 0$ by (P2); thus averaging the equality in (A.4) according to q_{t-1}^1 implies that

$$u\big(q_t^1, z_{t-1}^2\big) - u(z_{t-1}) > 0.$$

Adding the last two displayed inequalities completes the proof.

Step 3. Let $\pi_t := \pi_t^1 + \pi_t^2$ and $\alpha_t := \alpha_t^1 + \alpha_t^2$, and define

$$v_t := u\big(z_t^1, z_t^2\big) - \frac{1}{\rho}\pi_t - \sum_{r=t+1}^{\infty} \frac{M_2}{r^2},$$

where $M_2 := 2M + 2M_1/\rho$. Then

$$\mathrm{E}[v_t \mid h_{t-1}] \geqslant v_{t-1} + \frac{t-1}{t^2}\alpha_{t-1} \geqslant v_{t-1}, \tag{A.5}$$

and there exists a bounded random variable v such that $u(z_t^1, z_t^2) \to v$ as $t \to \infty$ a.s.

[26] Compare with (4.7) and with the computation of Lemma 2.2 in Hart and Mas-Colell (2001a).

[27] We use j for the other player (i.e., $j = 3 - i$).

Proof. We have

$$\mathrm{E}\big[t^2u(z_t^1,z_t^2)\mid h_{t-1}\big]=(t-1)^2u(z_{t-1}^1,z_{t-1}^2)+(t-1)u(q_t^1,z_{t-1}^2)+(t-1)u(z_{t-1}^1,q_t^2)+u(q_t^1,q_t^2)$$

and thus (recall the definition of α_{t-1}^i, and $|u(\cdot)|\leqslant M$)

$$\mathrm{E}\big[u(z_t^1,z_t^2)\mid h_{t-1}\big]\geqslant u(z_{t-1}^1,z_{t-1}^2)-\frac{t-1}{t^2}\pi_{t-1}+\frac{t-1}{t^2}\alpha_{t-1}-\frac{2M}{t^2}. \tag{A.6}$$

Using the inequality (A.2) of Step 1, $\pi_{t-1}\geqslant 0$, and $\alpha_{t-1}\geqslant 0$, we get

$$\begin{aligned}\mathrm{E}[v_t\mid h_{t-1}]&\geqslant u(z_{t-1}^1,z_{t-1}^2)-\frac{t-1}{t^2}\pi_{t-1}+\frac{t-1}{t^2}\alpha_{t-1}-\frac{2M}{t^2}-\frac{1}{\rho}(1-\rho/t)\pi_{t-1}-\frac{2M_1}{\rho t^2}-\sum_{r=t+1}^{\infty}\frac{M_2}{r^2}\\&\geqslant u(z_{t-1}^1,z_{t-1}^2)-\frac{1}{\rho}\pi_{t-1}-\sum_{r=t}^{\infty}\frac{M_2}{r^2}+\frac{t-1}{t^2}\alpha_{t-1}=v_{t-1}+\frac{t-1}{t^2}\alpha_{t-1}\geqslant v_{t-1}.\end{aligned}$$

Therefore $(v_t)_{t=1,2,\ldots}$ is a bounded submartingale, which implies that there exists a bounded random variable v such that $v_t\to v$ a.s., and so $u(z_t^1,z_t^2)\to v$ (since $\pi_t^i\to 0$ by Step 1).

Step 4. $\mathbf{1}_{\pi_t^1>0}(u(z_t)-u(z_t^1,z_t^2))\to 0$ as $t\to\infty$ a.s.

Proof. From (A.5) we get

$$w_T:=\sum_{t=1}^{T}\big(\mathrm{E}[v_{t+1}\mid h_t]-v_t\big)\geqslant\sum_{t=1}^{T}\frac{\beta_t}{t},\quad\text{where }\beta_{t-1}:=\tfrac{(t-1)^2}{t^2}\alpha_{t-1}\geqslant 0.$$

Thus $(w_T)_{T=1,2,\ldots}$ is a non-negative non-decreasing sequence, with $\sup_T\mathrm{E}(w_T)=\sup_T\mathrm{E}(v_{T+1})-\mathrm{E}(v_1)<\infty$ (the sequence v_t is bounded). Therefore a.s. $\lim w_T$ exists and is finite, which implies that

$$\sum_{t=1}^{\infty}\frac{\beta_t}{t}<\infty. \tag{A.7}$$

In addition, $|z_t(s)-z_{t-1}(s)|\leqslant 1/t$ for all $s\in S$, and therefore

$$|\beta_t-\beta_{t-1}|\leqslant\frac{M_3}{t}\quad\text{for some constant }M_3. \tag{A.8}$$

Lemma A.2. *Let $(\beta_t)_{t=1,2,\ldots}$ be a non-negative real sequence satisfying* (A.7) *and* (A.8)*. Then $\beta_t\to 0$ as $t\to\infty$.*

Proof. [28] Without loss of generality take $M_3=1$. Let $0<\varepsilon\leqslant 1$, and assume that $\beta_t\geqslant 2\varepsilon$ for some t. Then (A.8) yields, for all $t\leqslant r\leqslant t+\varepsilon t$,

$$\beta_r\geqslant\beta_t-\frac{1}{t+1}-\cdots-\frac{1}{r}\geqslant 2\varepsilon-\frac{r-t}{t}\geqslant 2\varepsilon-\varepsilon=\varepsilon,\quad\text{and thus}\quad\frac{\beta_r}{r}\geqslant\frac{\varepsilon}{(1+\varepsilon)t}\geqslant\frac{\varepsilon}{2t}.$$

Therefore

$$\sum_{t\leqslant r\leqslant t+\varepsilon t}\frac{\beta_r}{r}\geqslant\varepsilon t\frac{\varepsilon}{2t}=\frac{\varepsilon^2}{2}>0.$$

By (A.7), this implies that there can be at most finitely many t such that $\beta_t\geqslant 2\varepsilon$, so indeed $\beta_t\to 0$. □

Using Lemma A.2 shows that a.s. $\beta_t\to 0$ and so $\alpha_t\to 0$, which together with $\pi_t\to 0$ proves Step 4 (recall (A.3)).

[28] (A.7) implies that the Cesaro averages of the β_t converge to 0 (this is Kronecker's Lemma); together with (A.8), we obtain that the β_t themselves converge to 0.

Step 5. $\mathbf{1}_{\pi_t^1=0}(u(z_t) - u(z_t^1, z_t^2)) \to 0$ as $t \to \infty$ a.s.

Proof. Let $\gamma_t := \mathbf{1}_{\pi_t^1=0}$ be the indicator of the event $\pi_t^1 = 0$ and define $X_t := t(u(z_t) - u(z_t^1, z_t^2))$. Then $|X_t - X_{t-1}| \leqslant 4M$, and

$$\mathrm{E}[X_t \mid h_{t-1}, \gamma_{t-1} = 1] = (t-1)u(z_{t-1}) + u\big(z_{t-1}^1, q_{t-1}^2\big) - (t-1)u\big(z_{t-1}^1, z_{t-1}^2\big) - u\big(z_{t-1}^1, q_{t-1}^2\big) = X_{t-1}$$

(since $q_t^1 = z_{t-1}^1$ when $\gamma_{t-1} = 1$). Let $Y_t := \gamma_{t-1}(X_t - X_{t-1})$; then the Y_t are uniformly bounded martingale differences. Azuma's inequality[29] yields, for each $\varepsilon > 0$ and $r < t$,

$$\Pr\left[\sum_{\tau=r+1}^{t} Y_\tau > t\varepsilon\right] < \exp\left(-\frac{(t\varepsilon)^2}{2(4M)^2(t-r)}\right) \leqslant \exp(-\delta t)$$

where $\delta := \varepsilon^2/32M^2 > 0$, and thus

$$\Pr\left[\sum_{\tau=r+1}^{t} Y_\tau > t\varepsilon \text{ for some } r < t\right] < t\exp(-\delta t).$$

For each $t \geqslant 1$ define $R \equiv R(t)$ to be the maximal index $r < t$ such that $\gamma_r = 0$; if there is no such r, put $R(t) = 0$ and, for convenience, take $\gamma_0 \equiv 0$ and $X_0 \equiv 0$. Thus $\gamma_\tau = 1$ for $R + 1 \leqslant \tau \leqslant t - 1$ and $\gamma_R = 0$. Therefore $\gamma_{t-1}X_t - X_{R(t)+1} = \sum_{\tau=R(t)+2}^{t} Y_\tau$, and so

$$\Pr[\gamma_{t-1}X_t - X_{R(t)+1} > t\varepsilon] < t\exp(-\delta t). \tag{A.9}$$

The series $\sum_t t\exp(-\delta t)$ converges; therefore, by the Borel–Cantelli Lemma, the event of (A.9) happening for infinitely many t has probability 0. Thus a.s. $\gamma_{t-1}X_{t-1} - X_{R(t)} \leqslant \gamma_{t-1}X_t - X_{R(t)+1} + 8M \leqslant t\varepsilon + 8M$ for all t large enough (recall that $|X_t - X_{t-1}| \leqslant 4M$), which implies that

$$\overline{\lim}_{t\to\infty} \frac{1}{t}\gamma_t X_t \leqslant \overline{\lim}_{t\to\infty} \frac{1}{t} X_{R(t)} + \varepsilon.$$

Now either $R(t) \to \infty$, in which case $(1/t)X_{R(t)} \leqslant (1/R(t))X_{R(t)} \to 0$ by Step 4 since $\pi_{R(t)}^1 > 0$; or $R(t) = r_0$ for all $t \geqslant r_0$, in which case $(1/t)X_{R(t)} \leqslant (1/t)(4M)r_0 \to 0$. Thus $\mathbf{1}_{\pi_t^1=0}(u(z_t) - u(z_t^1, z_t^2)) = \gamma_t X_t/t \to 0$ a.s., as claimed.

Proof of Theorem A.1. Steps 4 and 5 show that $u(z_t)$ converges (a.s.) to the same (random) limit v of $u(z_t^1, z_t^2)$ (recall Step 3), which proves that any limit point of the sequence (z_t^1, z_t^2) is indeed a Nash equilibrium (see Lemma 4.1). □

Remark. The proof shows that in fact, with probability one, all limit points are Nash equilibria with the same payoff; that is, for almost every realization (i.e., infinite history) there exists an equilibrium payoff v such that $u(z_t^1, z_t^2)$—and also $u(z_t)$—converges to v.

Appendix B. Continuous-time Markov processes

In this appendix we prove a result on continuous-time Markov processes that we need in Section 5.

Proposition B.1. *For each $t \geqslant 1$, let $A(t)$ be a stochastic $m \times m$ matrix, and assume that there exists K such that*

$$\big\|A(t_2) - A(t_1)\big\| \leqslant K\frac{t_2 - t_1}{t_1} \quad \textit{for all } t_2 \geqslant t_1 \geqslant 1.$$

[29] Azuma's inequality is: $\Pr[\sum_{i=1}^{m} Y_i > \lambda] < \exp(-\lambda^2/(2K^2m))$, where the Y_i are martingale differences with $|Y_i| \leqslant K$; see Alon and Spencer (2000, Theorem 7.2.1).

Consider the differential system

$$\dot{x}(t) = x(t)\big(A(t) - I\big)$$

starting with some[30] $x(1) \in \Delta(m)$. *Then*

$$\dot{x}(t) \to 0.$$

The proof consists of considering first the case where $A(t) = A$ is independent of t (Proposition B.2), and then estimating the difference in the general case (Lemma B.3).

Proposition B.2. *There exists a universal constant c such that*

$$\big\| e^{t(A-I)}(A - I) \big\| \leqslant \frac{c}{\sqrt{t}}$$

for any stochastic matrix[31] A *and any* $t \geqslant 1$.

Proof. We have $e^{t(A-I)} = e^{-tI} e^{tA} = e^{-t} e^{tA}$ and

$$e^{tA}(A - I) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\big(A^{n+1} - A^n\big) = \sum_{n=0}^{\infty} \alpha_n A^n, \quad \text{where } \alpha_n := \frac{t^{n-1}}{(n-1)!} - \frac{t^n}{n!}$$

(put $t^{-1}/(-1)! = 0$). The matrix A^n is a stochastic matrix for all n; therefore $\|A^n\| = 1$, and thus

$$\big\| e^{tA}(A - I) \big\| \leqslant \sum_{n=0}^{\infty} |\alpha_n|.$$

Now $\alpha_n > 0$ for $n > t$ and $\alpha_n \leqslant 0$ for $n \leqslant t$, so $\sum_n |\alpha_n| = \sum_{n>t} \alpha_n - \sum_{n \leqslant t} \alpha_n$. Each one of the two sums is telescopic and reduces to[32] $t^r/r!$, where $r := \lfloor t \rfloor$ denotes the largest integer that is $\leqslant t$. Using Stirling's formula[33] $r! \sim \sqrt{2\pi r}\, r^r e^{-r}$ together with $t/r \to 1$ and $(t/r)^r \to e^{t-r}$ yields

$$\frac{t^r}{r!} \sim \frac{t^r e^r}{\sqrt{2\pi r}\, r^r} \sim \frac{e^t}{\sqrt{2\pi t}}.$$

Therefore

$$\overline{\lim}_{t\to\infty} \big\| e^{t(A-I)}(A - I) \big\| \sqrt{t} \leqslant \sqrt{\frac{2}{\pi}},$$

from which the result follows.[34] □

Remark. For each stochastic matrix A it can be shown that[35] $\|e^{t(A-I)}(A - I)\| = O(e^{\mu t})$, where $\mu < 0$ is given by[36] $\mu := \max\{\operatorname{Re}\lambda : \lambda \neq 0$ is an eigenvalue of $A - I\}$. However, this estimate—unlike the $O(t^{-1/2})$ of Proposition B.2—is *not* uniform in A and thus does not suffice.

---

[30] Recall that $\Delta(m)$ is the $(m-1)$-dimensional unit simplex in $\mathbb{R}^m$. Note that $x(1) \in \Delta(m)$ implies that $x(t) \in \Delta(m)$ for all $t \geqslant 1$.

[31] Of arbitrary size $m \times m$.

[32] They are equal since $\sum_n \alpha_n = 0$.

[33] $f(t) \sim g(t)$ means that $f(t)/g(t) \to 1$ as $t \to \infty$.

[34] Note that all estimates are uniform: They depend neither on A nor on the dimension m.

[35] $f(t) = O(g(t))$ means that there exists a constant c such that $|f(t)| \leqslant c|g(t)|$ for all t large enough.

[36] λ is an eigenvalue of $A - I$ if and only if $\lambda + 1$ is an eigenvalue of A. Thus $|\lambda + 1| \leqslant 1$, which implies that either $\lambda = 0$ or $\operatorname{Re}\lambda < 0$.

Lemma B.3. *For each $t \geqslant 1$, let $A(t), B(t)$ be stochastic $m \times m$ matrices, where the mappings $t \to A(t)$ and $t \to B(t)$ are continuous. Let $x(t)$ and $y(t)$ be, respectively, the solutions of the differential systems*

$$\dot{x}(t) = x(t)\big(A(t) - I\big) \quad \text{and} \quad \dot{y}(t) = y(t)\big(B(t) - I\big),$$

starting with some $x(1), y(1) \in \Delta(m)$. Then for all $t \geqslant 1$

$$\big\|x(t) - y(t)\big\| \leqslant \big\|x(1) - y(1)\big\| + \int_1^t \big\|A(\tau) - B(\tau)\big\|\,\mathrm{d}\tau.$$

Proof. Let $z(t) := \mathrm{e}^{t-1}x(t)$ and $w(t) := \mathrm{e}^{t-1}y(t)$, then $\dot{z}(t) = z(t)A(t)$ and $\dot{w}(t) = w(t)B(t)$. We have $\|\dot{w}(t)\| \leqslant \|w(t)\|\,\|B(t)\| = \|w(t)\|$, which implies that $\|w(t)\| \leqslant \mathrm{e}^{t-1}\|w(1)\| = \mathrm{e}^{t-1}$. Put $v(t) := z(t) - w(t)$; then

$$\begin{aligned}\big\|\dot{v}(t)\big\| &\leqslant \big\|\big(z(t) - w(t)\big)A(t)\big\| + \big\|w(t)\big(A(t) - B(t)\big)\big\| \\ &\leqslant \big\|z(t) - w(t)\big\|\big\|A(t)\big\| + \big\|w(t)\big\|\big\|A(t) - B(t)\big\| \leqslant \big\|v(t)\big\| + \mathrm{e}^{t-1}\delta(t),\end{aligned}$$

where $\delta(t) := \|A(t) - B(t)\|$. The solution of $\dot{\eta}(t) = \eta(t) + \mathrm{e}^{t-1}\delta(t)$ is

$$\eta(t) = \mathrm{e}^{t-1}\left(\eta(1) + \int_1^t \delta(\tau)\,\mathrm{d}\tau\right), \quad \text{so} \quad \big\|v(t)\big\| \leqslant \mathrm{e}^{t-1}\left(\big\|v(1)\big\| + \int_1^t \delta(\tau)\,\mathrm{d}\tau\right),$$

which, after dividing by e^{t-1}, is precisely our inequality. □

We can now prove our result.

Proof of Proposition B.1. Let $\alpha = 2/5$. Given T, put $T_0 := T - T^\alpha$. Let $y(T_0) = x(T_0)$ and $\dot{y}(t) = y(t)(A(T_0) - I)$ for $t \in [T_0, T]$. By Proposition B.2,

$$\big\|\dot{y}(T)\big\| \leqslant O\big((T - T_0)^{-1/2}\big) = O\big(T^{-\alpha/2}\big).$$

Now $\|A(t) - A(T_0)\| \leqslant K(t - T_0)/T_0 \leqslant KT^\alpha/(T - T^\alpha) = O(T^{\alpha-1})$ for all $t \in [T_0, T]$, and thus, by Lemma B.3, we get $\|x(T) - y(T)\| \leqslant (T - T_0)O(T^{\alpha-1}) = O(T^{2\alpha-1})$. Therefore

$$\begin{aligned}\big\|\dot{x}(T) - \dot{y}(T)\big\| &= \big\|x(T)\big(A(T) - I\big) - y(T)\big(A(T_0) - I\big)\big\| \\ &\leqslant \big\|x(T)\big\|\big\|A(T) - A(T_0)\big\| + \big\|x(T) - y(T)\big\|\big\|A(T_0) - I\big\| \\ &\leqslant O\big(T^{\alpha-1}\big) + O\big(T^{2\alpha-1}\big) = O\big(T^{2\alpha-1}\big).\end{aligned}$$

Adding the two estimates yields

$$\big\|\dot{x}(T)\big\| \leqslant O\big(T^{-\alpha/2}\big) + O\big(T^{2\alpha-1}\big) = O\big(T^{-1/5}\big)$$

(recall that $\alpha = 2/5$). □

Appendix C. Continuous-time approachability

We state and prove here the continuous-time analog of the Blackwell (1956) Approachability Theorem and its generalization in Hart and Mas-Colell (2001a, Section 2); all the notations follow the latter paper. The vector-payoff function is $A : S^i \times S^{-i} \to \mathbb{R}^m$, and we are given a convex closed set $C \subset \mathbb{R}^m$, which is *approachable*, i.e., for every $\lambda \in \mathbb{R}^m$ there exists $\sigma^i \in \Delta(S^i)$ such that

$$\lambda \cdot A\big(\sigma^i, s^{-i}\big) \leqslant w(\lambda) := \sup\{\lambda \cdot y\colon\ y \in C\} \quad \text{for all } s^{-i} \in S^{-i} \tag{C.1}$$

(see (2.1) there).

Let $P:\mathbb{R}^m \to \mathbb{R}$ be a C^1 function satisfying

$$\nabla P(x)\cdot x > w\big(\nabla P(x)\big) \quad \text{for all } x \notin C, \tag{C.2}$$

and also, without loss of generality,[37] $P(x) > 0$ for all $x \notin C$ and $P(x) = 0$ for all $x \in C$. We say that player i plays a *generalized approachability strategy* if the play $q^i(t) \in \Delta(S^i)$ of i at time t satisfies

$$\lambda(t)\cdot A\big(q(t), s^{-i}\big) \leqslant w\big(\lambda(t)\big) \quad \text{for all } s^{-i} \in S^{-i}, \tag{C.3}$$

where

$$\lambda(t) = \nabla P\big(A\big(z(t)\big)\big) \tag{C.4}$$

(such a $q^i(t)$ exists since C is approachable—see (C.1)). Note that the original Blackwell strategy corresponds to $P(x)$ being the squared Euclidean distance from x to the set C.

Theorem C.1. *Let $z(t)$ be a solution of* (2.1), (C.3) *and* (C.4). *Then $A(z(t)) \to C$ as $t \to \infty$.*

Proof. Rescale t so that $\dot{z} = q - z$. Denote $\pi(t) := P(A(z(t)))$. If $z(t) \notin C$, then

$$\dot{\pi} = \nabla P \cdot A(\dot{z}) = \lambda \cdot A\big(q^i, q^{-i}\big) - \lambda \cdot A(z) < w(\lambda) - w(\lambda) = 0$$

(we have used (C.4), (C.3), and (C.2)). Thus π is a strict Lyapunov function, and so $\pi \to 0$ as $t \to \infty$. □

References

Alon, N., Spencer, J.H., 2000. The Probabilistic Method, 2nd Edition. Wiley.

Benaïm, M., 1999. Dynamics of stochastic approximation algorithms. In: Azema, J., et al. (Eds.), Seminaire de Probabilites XXXIII. Lecture Notes in Math., Vol. 1709. Springer, pp. 1–68.

Benaïm, M., Weibull, J., 2003. Deterministic approximation of stochastic evolution in games. Econometrica 71, 873–903.

Blackwell, D., 1956. An analog of the minmax theorem for vector payoffs. Pacific J. Math. 6, 1–8.

Cahn, A., 2000. General procedures leading to correlated equilibria. The Hebrew Univ. of Jerusalem, Center for Rationality DP-216. Mimeo.

Fudenberg, D., Levine, D.K., 1998. Theory of Learning in Games. MIT Press.

Fudenberg, D., Levine, D.K., 1999. Conditional universal consistency. Games Econ. Behav. 29, 104–130.

Hannan, J., 1957. Approximation to Bayes risk in repeated play. In: Dresher, M., Tucker, A.W., Wolfe, P. (Eds.), Contributions to the Theory of Games, Vol. III. Ann. Math. Stud., Vol. 39. Princeton Univ. Press, pp. 97–139.

Hart, S., Mas-Colell, A., 2000. A simple adaptive procedure leading to correlated equilibrium. Econometrica 68, 1127–1150.

Hart, S., Mas-Colell, A., 2001a. A general class of adaptive strategies. J. Econ. Theory 98, 26–54.

Hart, S., Mas-Colell, A., 2001b. A reinforcement procedure leading to correlated equilibrium. In: Debreu, G., Neuefeind, W., Trockel, W. (Eds.), Economic Essays: A Festschrift for Werner Hildenbrand. Springer, pp. 181–200.

Hofbauer, J., 2000. From Nash and Brown to Maynard Smith: equilibria, dynamics and ESS. Selection 1, 81–88.

Hofbauer, J., Sandholm, W.H., 2002. On the global convergence of stochastic fictitious play. Econometrica 70, 2265–2294.

Rosenthal, R.W., 1973. A class of games possessing pure strategy Nash equilibria. Int. J. Game Theory 2, 65–67.

Sandholm, W.H., 2002. Potential dynamics and stable games. Univ. of Wisconsin. Mimeo.

[37] Cf. the Proof of Theorem 2.1 of Hart and Mas-Colell (2001a).