International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

Learning in Potential Games

Ermoliev, Y.M. and Flam, S.D.

IIASA Interim Report
June 1997

IIASA
International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-97-022 / June

Learning in Potential Games

Yuri M. Ermoliev (ermoliev@iiasa.ac.at)
Sjur Didrik Flåm (sjur.flaam@econ.uib.no)

Approved by
Gordon MacDonald (macdon@iiasa.ac.at)
Director, *IIASA*

About the Authors

Yuri M. Ermoliev is co-leader of the Risk, Modeling and Policy Project at IIASA, A-2361 Laxenburg, Austria.

Sjur Didrik Flåm is the corresponding author. He is at the Department of Economics, University of Bergen, N-5007 Norway.

Support from L. Meltzers høyskolefond, NFR project Quantec111039/410, and from IIASA is gratefully acknowledged.

Abstract

We consider repeated play of so-called potential games. Numerous modes of play are shown to yield Nash equilibrium in the long run. We point to procedures that can account for society-wide constraints concerning efficiency.

Key words: Potential games, repeated play, Nash equilibrium, learning, fictitious play, distributed computation.

Contents

Learning in Potential Games

Yuri M. Ermoliev (ermoliev@iiasa.ac.at)
Sjur Didrik Flåm (sjur.flaam@econ.uib.no)

1 Introduction

Game theory currently undergoes some re-orientation - away from strategic form, one-shot interaction between hyper-rational, omniscient players. Specifically, many recent studies allow repeated play of the same stage game, and accommodate agents enjoying merely bounded rationality, competence, or knowledge [12], [16]. A most natural question then becomes: *will the concerned parties learn Nash equilibrium over time?*

This question also motivates the present note. Our focus is on a special class consisting of so-called *potential games*, recently introduced and studied by Monderer and Shapley [13]. That class is admittedly narrow, but more important for economics - and more frequent in practice - than might first be believed [20]. So, part of our motivation is to show that manifold modes of play - associated with known methods of optimization - will all lead to equilibrium in games of this sort. To wit, we demonstrate below that quite flexible adjustments of strategies suffice to approach equilibrium over time. We believe that our convergence results, taken together, lend considerable relevance and justification to the Nash solution concept in potential games.

Additional motivation stems from the fact play must often be constrained so as to ensure long-term viability of the social interaction at hand. For example, in few-strategy games, say, prisoner's dilemma situations, there is often a free rider drift to low levels of efficiency. Then, concerns with equity, social security, financial sustainability, or outside competition may generate collective restrictions on how the game can be played [21], [23]. The approach used here can incorporate such concerns.

There is significant overlap between our modelling of repeated play and recent methods of distributed computation [2]. In fact, within the setting of potential games, this note can be read either as formalized fiction about iterated play, or as suggestions of algorithmic procedures. The reader will also see that quite often repeated play fits frames that make for a unified convergence analysis.

The note is organized as follows. Section 1 introduces the class of games under consideration. Since a potential, in the jargon of dynamic system theory, furnishes a Lyapunov function, it is convenient to collect, in Section 1, some immediate propositions about monotone convergence of potential values, and about existence of Nash equilibrium. Most adjustments of strategies advocated below share some important common features, - singled out in Section 3. The stage is then set for Section 4 to explore repeated play under the hypothesis that all agents come forth from time to time. That hypothesis is relaxed in Section 5, requiring merely an almost cyclic nature of active interventions. Section 6 gathers observations concerning what we naturally call gradient play. Section 7 indicates possibilities to offer agents almost unbounded freedom in the protocol governing the schedule of play. Section 8 concludes.

2 The Stage Game

A fixed non-cooperative stage game is played repeatedly at discrete time epochs (or stages). At each repetition players act sequentially, possibly in random order. Whenever a player comes forth he knows the most recent choices made by his rivals. In many scenarios described below he may, however, abstain from taking new actions over extended lapses of time.

There is finite, fixed set I of agents. Individual $i \in I$ always selects his strategy x_i within a prescribed *compact* set X_i. Let $x_{-i} := (x_j)_{j \neq i}$ denote the profile chosen by his rivals. In situation $x = (x_i, x_{-i})$ he obtains a payoff $\pi_i(x) = \pi_i(x_i, x_{-i})$ which is upper semicontinuous in x, and separately continuous in x$_{-i}$.

The game is said to have a *generalized ordinal potential* $P : \Pi_{i \in I} X_i \to \mathcal{R}$ if for every two strategy profiles (x_i^{+1}, x_{-i}), $(x_i, x_{-i}) \in X := \Pi_{i \in I} X_i$ we have

$$\pi_i(x_i^{+1}, x_{-i}) > \pi_i(x_i, x_{-i}) \Longrightarrow P(x_i^{+1}, x_{-i}) > P(x_i, x_{-i}). \tag{1}$$

(The superscript +1 used here and below might first appear a somewhat strange notation. The reader will, however, soon find it natural.) A generalized potential is declared *squeezing* if

$$P(x_i^{+1}, x_{-i}) - P(x_i, x_{-i}) \downarrow 0 \Rightarrow \max\left\{0, \pi_i(x_i^{+1}, x_{-i}) - \pi_i(x_i, x_{-i})\right\} \to 0. \tag{2}$$

Squeezing obtains if

$$\pi_i(x_i^{+1}, x_{-i}) > \pi_i(x_i, x_{-i}) \Leftrightarrow P(x_i^{+1}, x_{-i}) > P(x_i, x_{-i}),$$

P then being called an *ordinal potential.* More generally, (2) holds when P has the following *sandwich property:* For all $i \in I$ there exists a positive number w_i such that

$$\pi_i(x_i^{+1}, x_{-i}) \geq \pi_i(x_i, x_{-i}) \Rightarrow w_i \left\{P(x_i^{+1}, x_{-i}) - P(x_i, x_{-i})\right\} \geq \pi_i(x_i^{+1}, x_{-i}) - \pi_i(x_i, x_{-i}). \tag{3}$$

In particular, (3), whence (2), is satisfied if

$$\pi_i(x_i^{+1}, x_{-i}) - \pi_i(x_i, x_{-i}) = w_i \left\{P(x_i^{+1}, x_{-i}) - P(x_i, x_{-i})\right\},$$

the ordinal potential P then being named *weighted*, and *exact* if all $w_i = 1$.

Potentials of all sorts are henceforth assumed to be *continuous upwards* in the sense that $\lim_{t \longrightarrow \infty} P(x^t) = P(x)$ whenever $\lim_{t \to \infty} x^t = x$ and $\{P(x^t)\}$ is nondecreasing, bounded above. (Note that such functions P need be neither upper nor lower semicontinuous.) We shall time and again use the following straightforward result:

Proposition 1 *(Convergence upwards) Let P be a generalized ordinal potential. Whenever repeated play gives a non-decreasing bounded sequence $\{P(x^t)\}$ it holds for every cluster point $x = \lim_{t\in T} x^t$ of $\{x^t\}$ that $P(x) = \lim_{t\in T} P(x^t) = \lim_{t\to\infty} P(x^t)$.* $\square$

Here and in the sequel we simply write $\lim_{t\in T} x^t$ for the more complete and correct expression $\lim_{t\in T, t\to\infty} x^t$.

We remark in passing that existence of Nash equilibrium is ensured under weak conditions:

Proposition 2 *(Existence of solutions) Let P be an upper semicontinuous generalized ordinal potential. Then there exists at least one Nash equilibrium.*

Proof. P being upper semicontinuous on the compact set X, guarantees that $\arg\max P$ is non-empty.. Evidently, every point in this set must be a Nash equilibrium. $\square$

One might be tempted to think that potential games enjoy distinguished welfare properties. This is, however, far from so. Indeed, the prisoners' dilemma and some Cournot oligopolies are potential games that yield inefficient equilibrium outcomes [13]. In those and many other cases, including congestion games [20], the hidden potential should not be seen as a welfare indicator. Rather, in terms of optimization theory, it constitutes a most natural merit function, monitoring the drift towards Nash equilibrium. This will be abundantly illustrated in the sequel.

3 Adjustments of Strategies

It is desirable and expedient to focus briefly on common features of many modes of repeated play.

Definition 1 *By an adjustment rule of individual i we understand a closed correspondence $\mathfrak{A}_i : X \rightsquigarrow X_i$.*

The interpretation is that if given the chance to reconsider his choice, individual i would in situation x select a new strategy $x_i \in \mathfrak{A}_i(x)$. Frequently it is convenient to identify an adjustment rule $\mathfrak{A}_i$ with the closed self-correspondence

$$A_i(x) := (\mathfrak{A}_i(x), x_{-i})$$

on X, - and we shall not hesitate to do so. As said, players will act sequentially at every stage, - in constrained or quite free order. To codify such freedom succinctly we write, whenever $\mathcal{O} = (i_1, ..., i_k)$ is an ordered subset of individuals from I,

$$A_{\mathcal{O}} := A_{i_k} \circ \ldots \circ A_{i_1}$$

for the resulting composite correspondence. Note that $A_{\mathcal{O}}$ so defined is closed. Nothing precludes, in principle, that adjustment rules and orders change in the course of play. Thus interaction at time t generally involves an order $\mathcal{O}_t$ in which the active players then enter the scene, each among them applying a personal, time-specific adjustment rule.

Definition 2 *(Monotonicity) A self-correspondence A on X is declared non-decreasing in potential values if*

$$x^{+1} \in A(x) \Longrightarrow P(x^{+1}) \geq P(x), \tag{4}$$

and improving if, in addition, the inequality in (4) is strict whenever $x \notin A(x)$. Correspondingly, a set of individual adjustment rules $\mathfrak{A}_i$, $i \in I$, is said to be potential non-decreasing (improving) if for every order $\mathcal{O}$ of applying $A_i(x) := (\mathfrak{A}_i(x), x_{-i})$, the composite correspondence $A_{\mathcal{O}}$ is of the same sort.

In (4) and elsewhere the notation x^{+1} is intended to emphasize dynamics, i.e., the "state" x^{+1} comes up +1 time step after x. As an example, suppose all adjustment rules increase own payoff, this meaning that for all i we have

$$x_i^{+1} \in \mathfrak{A}_i(x) \Longrightarrow \pi_i(x_i^{+1}, x_{-i}) \geq \pi_i(x_i, x_{-i}),$$

with strict inequality whenever $x_i \notin \arg\max \pi_i(\cdot, x_{-i})$. Then evidently, employing such rules will improve the value of a generalized ordinal potential.

Whatever rules are used, they should, at least ideally, not cause a premature halt at disequilibrium. It is convenient to express this property compactly:

Definition 3 *(Nash fixed points) A self-correspondence A on X is declared acceptable if all its fixed points are Nash equilibria. Similarly, a set of individual adjustment rules $\mathfrak{A}_i$, $i \in I$, is named acceptable if for every order $\mathcal{O}$ of applying <u>all</u> $A_i(x) := (\mathfrak{A}_i(x), x_{-i}), i \in I$, the composite correspondence $A_{\mathcal{O}}$ is of the same sort.*

Definition 4 *(Maximizing fixed points) A self-correspondence A on X is said to be adapted to the potential P if all its fixed points belong to* $\arg\max P$. *Correspondingly, a set of individual adjustment rules $\mathfrak{A}_i$, $i \in I$, is said to be adapted to P if for every order $\mathcal{O}$ of applying <u>all</u> $A_i(x) := (\mathfrak{A}_i(x), x_{-i}), i \in I$, the composite correspondence $A_{\mathcal{O}}$ is of the said sort.*

4 Repetitive Play

We begin by considering particular modes of repeated play, called *repetitive,* requiring *all* agents to actively use *fixed* adjustment rules at every stage.

Theorem 1 *(Convergence under a stationary mode of play) Suppose the game has a generalized ordinal potential. Also suppose that all agents always actively use fixed, improving, acceptable adjustment rules. Then repeated play, proceeding in prescribed or free order at any stage, will cluster to Nash equilibrium.*

Proof. The demonstration parallels that of Zangwill's first convergence theorem [24]. Let $x = \lim_{t \in T} x^t$ be a cluster point (accumulation point) of the sequence $\{x^t\}$ generated by repeated play. Assume, without loss of generality, that

$$\lim_{t \in T} x^{t+1} =: x^{+1},$$

and that the same order $\mathcal{O}$ of play is used for all $t \in T$. Then

$$x^{+1} \in A_{\mathcal{O}}(x).$$

If x is not a fixed point of the improving correspondence $A_{\mathcal{O}}$, then $P(x^{+1}) > P(x)$. However, by Proposition 1 it holds that $P(x^{+1}) = P(x)$. So, to avoid this contradiction, x must indeed be a fixed point of the acceptable $A_{\mathcal{O}}$, whence a Nash equilibrium. □

Note, by the way, that this result generalizes and simplifies Proposition 2.7.1 in Bertsekas [3]. Next we provide two immediate applications:

Corollary 1 *(Convergence of best reply) Suppose the game has a generalized ordinal potential. Suppose also that all agents invariably use best reply*

$$\mathfrak{A}_i(x) := \arg\max \pi_i(\cdot, x_{-i}) \tag{5}$$

as their adjustment rules. Then iterative best reply, proceeding in prescribed or free order at any stage, will cluster to Nash equilibrium. □

Corollary 2 *(Convergence of averaged best reply) Suppose the game has a generalized ordinal potential and that all decision sets X_i are convex. Suppose also that every agent i has concave payoff $\pi_i(x_i, x_{-i})$ in own decision x_i, and that, for fixed $\lambda_i \in \left]0, 1\right]$, he invariably uses averaged best reply*

$$\mathfrak{A}_i(x) := (1 - \lambda_i)x_i + \lambda_i \arg\max \pi_i(\cdot, x_{-i}) \tag{6}$$

as his adjustment rule. Then repeated play, in prescribed or free order at any stage, will cluster to Nash equilibrium. □

We proceed next to relax the double assumption above that players use *fixed* adjustment rules and intervene at *every* stage.

Theorem 2 *(Convergence under spacer steps) Assume the game has a generalized ordinal potential. Assume also that agents always cause the potential to be non-decreasing, and that infinitely often all players implement time-invariant, improving, adapted adjustment rules. Then repeated play, in prescribed or free order at any stage, will cluster to Nash equilibrium belonging to* $\arg\max P$.

Proof. Consider the sequence $T \subseteq \{0, 1, ...\}$ of times t at which the same acceptable adjustment rules come into application. Without loss of generality assume that

$$\lim_{t\in T} x^t =: x, \quad \lim_{t\in T} x^{t+1} =: x^{+1},$$

and that the same order $\mathcal{O}$ of play is used at all times $t \in T$. Consequently,

$$x^{+1} \in A_{\mathcal{O}}(x) \text{ and } \quad P(x^{+1}) = P(x).$$

If x is not a fixed point of the improving $A_{\mathcal{O}}$, then $P(x^{+1}) > P(x)$, - a contradiction. So, x is indeed a fixed point of the adapted $A_{\mathcal{O}}$, whence a Nash equilibrium belonging to $\arg\max P$. Since for every cluster point $\overline{x}$ of the generated sequence $\{x^t\}$ it holds that $P(\overline{x}) = P(x) = \max P$, all these points must belong to $\arg\max P$. □

Remark. (Inhomogeneous adjustments) For any sequence $\{A^t\}$ of self-correspondences on X define a "cluster" correspondence, denoted $\limsup_{t\to\infty} A^t$, by

$$x^{+1} \in \limsup_{t\to\infty} A^t(x)$$

iff there exists a subsequence T and associated points $x^{t+1} \in A^t(x^t)$, $\forall t \in T$, such that

$$\lim_{t\in T} x^{t+1} = x^{+1} \quad \text{and} \quad \lim_{t\in T} x^t = x.$$

Evidently, Theorem 1 continues to hold with time-variant, nondecreasing adjustment rules provided that along every subsequence T

$$\limsup_{t\in T} A^t \subseteq A \tag{7}$$

for some improving, acceptable correspondence A. Likewise, Theorem 2 remains valid with changing, nondecreasing adjustment rules granted that along some subsequence T condition (7) holds with an adapted A. □

Example 1 *(Almost best reply) In (5) one might rather use*

$$\mathfrak{A}_i^t(x) := \varepsilon_{it}\text{-}\arg\max \pi_i(\cdot, x_{-i})$$

with $\varepsilon_{it} \to 0, \forall i$. *The assertion in Corollary 1 would still hold.* □

Example 2 *(Moving averages) Instead of (6) we could very well apply*

$$\mathfrak{A}_i^t(x) := (1-\lambda_{it})x_i + \lambda_{it} \arg\max \pi_i(\cdot, x_{-i})$$

with $0 < \lambda_i \le \lambda_{it} \le 1, \forall i, t$. *This would not alter the conclusion of Corollary 2.* □

In the same vein one may introduce *variational (or so-called gamma-) convergence* of payoff functions [6]. Doing so would allow players - in the "interim" - to use surrogate or approximate objectives. The latter must converge though, in appropriate sense, to the true payoff functions [4], [5].

5 Almost Cyclic Play

It is natural to relax the preceding hypothesis that *all* players act at infinitely many stages. So, here we shall allow each of them to be inactive for substantial lapses of time, even in non-coordinated manners, provided, of course, they intervene every so often.

Definition 5 *(Almost cyclic interventions) Let* $I_t \subseteq I$ *denote the set of players who indeed take actions at time (stage)* t. *Play is said to be almost cyclic if there exists some finite integer period* $p \ge 1$ *such that*

$$I_t \bigcup \ldots \bigcup I_{t+p-1} = I \text{ for all } t. \tag{8}$$

Typically, convergence of play amounts to have $x^{t+1} \approx x^t$ for all large t. To codify this phenomenon we introduce the concept of *asymptotic regularity*:

Definition 6 *A mode of repeated play is called asymptotically regular if for every generated sequence* $\{x^t\}$ *we have*

$$\left\|x^{t+1} - x^t\right\| \to 0. \tag{9}$$

It is from now on tacitly understood that all decision sets X_i are contained in real Hilbert spaces, and that the norm $\|\cdot\|$ on the product set X is of the L_2-type, i.e.,

$$\left\|x^{+1} - x\right\|^2 = \sum_i \left\|x_i^{+1} - x_i\right\|_i^2. \tag{10}$$

For convenience we shall henceforth not discriminate notationally between norms (and between inner products) in different spaces.

Theorem 3 *(Convergence of almost cyclic play) Suppose the game has a generalized ordinal potential. Also suppose that play is almost cyclic, asymptotically regular, and involves potential improving, fixed adjustment rules. Then repeated play, in fixed or random order at any stage, will cluster to Nash equilibrium.*

Proof. Suppose some cluster point $x = \lim_{t\in T} x^t$ is *not* a Nash equilibrium. Then there is at least one individual i, henceforth singled out, who would regret the choice x_i when confronted with the profile x_{-i} of his rivals. For each $t \in T$, find a time $\tau(t) \in \{t, ..., t+p-1\}$ at which the considered player i does indeed seize the opportunity to act. By asymptotic regularity

$$\lim_{t\in T} x^{\tau(t)+1} = \lim_{t\in T} x^{\tau(t)} = \lim_{t\in T} x^t = x.$$

Assume, without loss of generality, that the same ordered set $\mathcal{O}$ of players comes forth at all times $\tau(t)$. (If not, pass to a subsequence.) Now $x^{\tau(t)+1} \in A_{\mathcal{O}}(x^{\tau(t)})$, $\forall$ $t \in T$, implies $x \in A_{\mathcal{O}}(x)$. However, individual i belongs to $\mathcal{O}$. Thus he, or some other dissatisfied agent, will improve strictly the potential at x, that is,

$$x^{+1} \in A_{\mathcal{O}}(x) \Longrightarrow P(x^{+1}) > P(x),$$

contradicting the fact that $x \in A_{\mathcal{O}}(x)$. □

Games with adjustment cost. To illustrate the preceding theorem, suppose agent i incurs additional cost

$$a_i(x_i^{+1}, x_i) \geq 0 \tag{11}$$

if and when he opts to change his last decision x_i to x_i^{+1} next time he acts. The idea here is that such adjustment cost cause players to move with some inertia. Intuitively, they will avoid making large deviations because such undertakings are too expensive. In any case, we naturally posit that

$$a_i(x_i, x_i) = 0 \text{ for all } x_i \in X_i; \tag{12}$$

$$a_i(x_i^{t+1}, x_i^t) \to 0 \Longrightarrow \left\|x_i^{t+1} - x_i^t\right\| \to 0; \tag{13}$$

$$x_i \in \arg\max\{\pi_i(\cdot, x_{-i}) - a_i(\cdot, x_i)\} \Rightarrow x_i \in \arg\max \pi_i(\cdot, x_{-i}). \tag{14}$$

Example 3 *(Quadratic adjustment cost) A natural candidate is*

$$a_i(x_i^{+1}, x_i) = \alpha_i \left\| x_i^{+1} - x_i \right\|^2$$

for some positive number α_i*. Then (11), (12), and (13) are immediate, whereas (14) holds provided* $\pi_i(\cdot, x_{-i})$ *is concave and differentiable (in the generalized sense of convex analysis [19]) on* X_i*. Indeed,*

$$x_i \in \arg\max \{\pi_i(\cdot, x_{-i}) - a_i(\cdot, x_i)\} \Rightarrow 0 \in \partial \{\pi_i(\cdot, x_{-i}) - a_i(\cdot, x_i)\}(x_i) \Rightarrow \\ 0 \in \partial \{\pi_i(\cdot, x_{-i})\}(x_i) \Rightarrow x_i \in \arg\max \pi_i(\cdot, x_{-i}). \ \square$$

Example 4 *(Bregman adjustment cost) More generally, let the real-valued function* ψ_i *be strictly convex and continuously differentiable in a neighborhood of the convex set* X_i*, and define*

$$a_i(x_i^{+1}, x_i) := \psi_i(x_i^{+1}) - \psi_i(x_i) - \left\langle \nabla \psi_i(x_i), x_i^{+1} - x_i \right\rangle \tag{15}$$

to be a so-called Bregman distance (see [22] and references therein). Again, (11) and (12) follow immediately, and (13) results from the strict convexity of ψ_i*. Finally, if* $\pi_i(\cdot, x_{-i})$ *is concave and generalized differentiable on* X_i*, then (14) is derived as above.* $\square$

Theorem 4 *(Convergence of best reply with adjustment cost) Suppose the game has a bounded squeezing potential (2), and is played almost cyclically with adjustment rules*

$$\mathfrak{A}_i(x) := \arg\max \{\pi_i(\cdot, x_{-i}) - a_i(\cdot, x_i)\}. \tag{16}$$

Then repeated play, in any order, clusters to Nash equilibrium.

Proof. Consider a fixed agent i. Let $x^{i,t}$ denote the profile which prevails just before he acts at time t. Then, via (12),

$$\pi_i(x_i^{t+1}, x_{-i}^{i,t}) - a_i(x_i^{t+1}, x_i^t) \geq \pi_i(x_i^t, x_{-i}^{i,t}) - a_i(x_i^t, x_i^t) = \pi_i(x_i^t, x_{-i}^{i,t}),$$

i.e.,

$$\pi_i(x_i^{t+1}, x_{-i}^{i,t}) - \pi_i(x^{i,t}) \geq a_i(x_i^{t+1}, x_i^t) \geq 0, \tag{17}$$

so that, by (1), potential values must be non-decreasing, and $\{P(x^t)\}$ converges. As a result, play is *asymptotically regular* (9). In fact,

$$\begin{aligned} 0 &\leftarrow P(x^{t+1}) - P(x^t) \geq P(x_i^{t+1}, x_{-i}^{i,t}) - P(x^{i,t}) \geq 0 \\ &\Rightarrow \pi_i(x_i^{t+1}, x_{-i}^{i,t}) - \pi_i(x^{i,t}) \to 0 \text{ by squeezing (2)} \\ &\Rightarrow a_i(x_i^{t+1}, x_i^t) \to 0 \text{ via (17)} \\ &\Rightarrow \left\| x_i^{t+1} - x_i^t \right\| \to 0 \text{ by (13).} \end{aligned}$$

Thus (10) ensures $\|x^{t+1} - x^t\| \to 0$. Finally, appealing to Theorem 3, the conclusion obtains. $\square$

6 Gradient Play

In this section we shall suppose that each payoff function $\pi_i(\cdot, x_{-i})$ is continuously differentiable on X_i, this set being convex compact and contained a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|\cdot\|$. We begin with

6.1 Gradient Projection Play

As is well known, the gradient (*marginal payoff*)

$$\nabla_i \pi_i(x) := \frac{\partial \pi_i(x_i, x_{-i})}{\partial x_i}$$

offers the direction of *steepest payoff ascent.* An adjustment along this direction may, however, lead out of the feasible set X_i. Therefore, letting $[\cdot]_i^+$ denote the *orthogonal projection* onto X_i, it is natural to consider adjustment rules

$$\mathfrak{A}_i(x) := [x_i + \rho_i \nabla_i \pi_i(x)]_i^+ , \tag{18}$$

called *gradient projection.* Here, of course, the *stepsize* ρ_i is positive. A strategy profile $x \in X$ is called a *stationary point of player* i iff $\langle \nabla_i \pi_i(x), \tilde{x}_i - x_i\rangle \leq 0$ for all $\tilde{x}_i \in X_i$.

Theorem 5 *(Convergence of gradient projection play) Suppose the game*

** has a bounded squeezing potential, and*

** is played almost cyclically with gradient projection rules, using stepsizes* $\rho_{it} \in \left[\varepsilon_i, \frac{2}{K_i} - \varepsilon_i\right]$ *where* $0 < \varepsilon_i < \frac{2}{K_i}$, *and* $K_i > 0$ *is a Lipschitz constant of marginal payoff, i.e.,*

$$\|\nabla_i \pi_i(x_i, x_{-i}) - \nabla_i \pi_i(\tilde{x}_i, x_{-i})\| \leq K_i \|x_i - \tilde{x}_i\| . \tag{19}$$

Then repeated play, in any order, clusters to stationary points. If all payoff functions π_i *are concave in own decision* x_i*, such points must be Nash equilibria. Given an exact potential* P*, any limit point* x *will satisfy*

$$P'(x; \tilde{x} - x) \leq 0 \text{ for all } \tilde{x} \in X. \tag{20}$$

If, moreover, P *is concave, then play accumulates to* $\arg\max P$.

Proof. For notational convenience write

$$x_i^{+1} := \mathfrak{A}_i(x) = [x_i + \rho_i \nabla_i \pi_i(x)]_i^+ ,$$

where the stepsize $\rho_i > 0$ temporarily is regarded as fixed. This last relation implies that

$$\left\langle x_i - x_i^{+1}, x_i + \rho_i \nabla_i \pi_i(x)_i - x_i^{+1}\right\rangle \leq 0,$$

from which we derive

$$\left\|x_i^{+1} - x_i\right\|^2 \leq \rho_i \left\langle \nabla_i \pi_i(\, x), x_i^{+1} - x_i\right\rangle . \tag{21}$$

Consider now a player i who faces a strategy profile x. Letting

$$f(s) := \pi_i(x_i + s(x_i^{+1} - x_i), x_{-i}), \quad s \in [0,1],$$

we get

$$\begin{aligned}
\pi_i(x_i^{+1}, x_{-i}) - \pi_i(x_i, x_{-i}) &= f(1) - f(0) = \int_0^1 f'(s)ds \\
&= f'(0) + \int_0^1 [f'(s) - f'(0)]\, ds \\
&\geq f'(0) - \int_0^1 |f'(s) - f'(0)|\, ds \\
&\geq f'(0) - \frac{K_i}{2} \left\| x_i^{+1} - x_i \right\|^2 \text{ by (19)} \\
&= \left\langle \nabla_i \pi_i(x), x_i^{+1} - x_i \right\rangle - \frac{K_i}{2} \left\| x_i^{+1} - x_i \right\|^2 \\
&\geq (\frac{1}{\rho_i} - \frac{K_i}{2}) \left\| x_i^{+1} - x_i \right\|^2 \geq 0 \text{ by (21).}
\end{aligned}$$

As before, denote by $x^{i,t}$ the situation confronting individual i when he acts at time t. The above string of inequalities implies first, via (1), that $\{P(x^t)\}$ is non-decreasing, hence convergent. Second, this same string yields for all i,

$$\begin{aligned}
0 &\leftarrow P(x^{t+1}) - P(x^t) \geq P(x_i^{t+1}, x_{-i}^{i,t}) - P(x^{i,t}) \geq 0 \\
&\Rightarrow \pi_i(x_i^{t+1}, x_{-i}^{i,t}) - \pi_i(x_i^t, x_{-i}^{i,t}) \to 0 \text{ by squeezing (2)} \\
&\Rightarrow \left\| x_i^{t+1} - x_i^t \right\| \to 0 \text{ by the above string,}
\end{aligned}$$

whence $\|x^{t+1} - x^t\| \to 0$. That is, the process $\{x^t\}$ is asymptotically regular (9). Let $x = \lim_{t \in T} x^t$ be any cluster point. Since by asymptotic regularity $x = \lim_{t\in T} x^{t+1}$, and since play is almost cyclic (8), the continuity of marginal payoff $\nabla_i \pi_i(\cdot)$ tells that

$$x_i = [x_i + \rho_i \nabla_i \pi_i(x)]_i^+ ,$$

for some subsequence such that $\rho_{it} \to \rho_i > 0$. Hence

$$\langle \nabla_i \pi_i(x), \tilde{x}_i - x_i \rangle \leq 0 \text{ for all } \tilde{x}_i \in X, \text{ and all } i. \tag{22}$$

When individual payoffs are concave in own decision, this suffices for x to be a Nash equilibrium. With an exact potential P we have

$$\nabla_i \pi_i(x) = \frac{\partial P(x)}{\partial x_i} \text{ for all } i,$$

and then the variational inequalities (22) yield (20). □

It may very well happen that some player i, during a transient phase, make steps that exceed the bound $\frac{2}{K_i}$. To ensure convergence all the same it is natural to play with *decreasing stepsizes* $\rho_{it} \to 0$. Section 7 returns to this issue in a slightly different setting.

6.2 Conditional Gradient Play

We continue to focus on gradient modes of play. Our objective right here is to get around the projection operator in (18) which may be hard to execute. Towards that objective we assume now that each active agent first looks for a direction in which he can better himself and thereafter proceeds along that line.

By a *feasible direction* of player i at $x \in X$ we understand a vector $d_i \neq 0$ such that $(x_i + \rho_i d_i, x_{-i}) \in X$ for all $\rho_i > 0$ sufficiently small. We posit here that agents always move in feasible directions. Recall that x is called a *stationary point of player* i iff $\langle \nabla_i \pi_i(x), d_i \rangle \leq 0$ for all feasible directions d_i at x. It is taken for granted that unless x is stationary for individual i, then

$$\langle \nabla_i \pi_i(x), d_i \rangle > 0, \tag{23}$$

for the chosen direction d_i, and

$$\pi_i(x_i + \rho_i d_i, x_{-i}) > \pi_i(x_i, x_{-i}) \tag{24}$$

for the implemented stepsize ρ_i.

Definition 7 *A direction sequence* $\{d_i^t\}$ *of player* i *is declared gradient related to* $\{x^t\}$ *if for any subsequence* $\{x^t\}_{t\in T}$ *converging to a point which is non-stationary for this player, the corresponding sequence* $\{d_i^t\}_{t\in T}$ *is bounded and satisfies*

$$\liminf_{t\in T} \left\langle \nabla_i \pi_i(x^t), d_i^t \right\rangle > 0. \tag{25}$$

Definition 8 *(Armijo stepsizes [1]) Player* $i \in I$ *is said to apply an Armijo rule at* $x \in X$ *if his stepsize* ρ_i *along any feasible direction* d_i*, for given scalars* $\beta_i, \sigma_i \in \left]0,1\right[$, $s_i > 0$, *equals the smallest number* $\beta_i^k s_i$, $k = 0, 1, ...$ *which satisfies*

$$\pi_i(x_i + \rho_i d_i, x_{-i}) - \pi_i(x_i, x_{-i}) \geq \sigma_i \rho_i \langle \nabla_i \pi_i(x), d_i \rangle . \tag{26}$$

Theorem 6 *(Convergence of feasible direction play) Suppose*

* *the game has a bounded squeezing potential,*

* *each* $\nabla_i \pi_i(x)$ *exists and is continuous on an Euclidean space, and*

$$\pi_i(x_i^{t+1}, x_{-i}^{i,t}) - \pi_i(x_i^t, x_{-i}^{i,t}) \downarrow 0 \Rightarrow \left\| x_i^{t+1} - x_i^t \right\| \to 0; \tag{27}$$

* *directions are gradient related, and players either use Armijo rules, or optimize their steplengths over full or limited range.*

Then repeated play, in any order, clusters to stationary points. If moreover, the potential P *is exact and jointly concave, then play will accumulate to maximizing points.*

Proof. Suppose first that all agents invariably apply Armijo stepsizes. Let $x^{i,t}$ denote the situation seen by agent i when he enters the scene at time t. By (23) and (26)

$$\pi_i(x_i^{t+1}, x_{-i}^{i,t}) - \pi_i(x^{i,t}) \geq \sigma_i \rho_{it} \left\langle \nabla_i \pi_i(x^{i,t}), d_i^t \right\rangle \geq 0.$$

Note that play stops at stationary points. Until then the potential increases steadily by (24), and converges by Proposition 1. It follows from the convergence of $\{P(x^t)\}$ and the squeezing (2) that $\pi_i(x_i^{t+1}, x_{-i}^{i,t}) - \pi_i(x^{i,t}) \to 0$. Hence,

$$\rho_{it} \left\langle \nabla_i \pi_i(x^{i,t}), d_i^t \right\rangle \to 0,$$

and, via (27), $\|x^{i,t} - x^t\| \to 0$.

Assume, for the sake of the argument, that some subsequential limit $x = \lim_{t\in T} x^t$ is *not* a stationary point of some players. Since directions are gradient related and $\|x^{i,t} - x^t\| \to 0$, we get via (25) that

$$\liminf_{t\in T} \left\langle \nabla_i \pi_i(x^{i,t}), d_i^t \right\rangle > 0,$$

whence

$$\lim_{t\in T} \rho_{it} = 0 \tag{28}$$

for all those agents who do not face stationarity in the cluster point x. Consider henceforth only those agents. (28) implies, by the very definition of the Armijo rule, that *no* initially proposed stepsize s_i were implemented from remote enough time $t \in T$ onwards. In other words, each s_i will eventually be reduced at least once by the factor β_i. Thus,

$$\pi_i(x_i^t + \frac{\rho_{it}}{\beta_i} d_i^t, x_{-i}^{i,t}) - \pi_i(x^{i,t}) < \sigma_i \frac{\rho_{it}}{\beta_i} \left\langle \nabla_i \pi_i(x^{i,t}), d_i^t \right\rangle.$$

Introduce

$$\overline{d}_i^t := \frac{d_i^t}{\|d_i^t\|}, \quad \overline{\rho}_{it} := \frac{\rho_{it}\,\|d_i^t\|}{\beta_i},$$

and use the mean value theorem to rewrite the preceding inequality as

$$\left\langle \nabla_i \pi_i(x_i^t + \widetilde{\rho}_{it} \overline{d}_i^t, x_{-i}^{i,t}), \overline{d}_i^t \right\rangle = \frac{\pi_i(x_i^t + \overline{\rho}_{it} \overline{d}_i^t, x_{-i}^{i,t}) - \pi_i(x^{i,t})}{\overline{\rho}_{it}} < \sigma_i \left\langle \nabla_i \pi_i(x^{i,t}), \overline{d}_i^t \right\rangle$$

for some number $\widetilde{\rho}_{it} \in [0, \overline{\rho}_{it}]$. The boundedness of $\{d_i^t\}_{t\in T}$ and (28) allow us to assert that $\lim_{t\in T} \overline{\rho}_{it} = 0$, whence $\lim_{t\in T} \widetilde{\rho}_{it} = 0$. Assume, without loss of generality, that $\lim_{t\in T} \overline{d}_i^t = \overline{d}_i$. Then, passing to the limit along $t \in T$ in the last string we get

$$\left\langle \nabla_i \pi_i(x), \overline{d}_i \right\rangle \le \sigma_i \left\langle \nabla_i \pi_i(x), \overline{d}_i \right\rangle,$$

saying, since $\sigma_i \in \left]0,1\right[$, that

$$\left\langle \nabla_i \pi_i(x), \overline{d}_i \right\rangle \le 0.$$

On the other hand, via (25),

$$\left\langle \nabla_i \pi_i(x), \overline{d}_i \right\rangle > 0.$$

This contradiction entails that x must be stationary under the Armijo rule. The other rules generate as much value added to payoffs at any stage as does the Armijo regime. Repeating the above argument as if an Armijo step would have been taken at any stage we get the same result as before. □

Remarks. The hypothesis (27) is fulfilled when

$$\pi_i(x_i^{+1}, x_{-i}) > \pi_i(x_i, x_{-i}) \Rightarrow \pi_i(x_i^{+1}, x_{-i}) - \pi_i(x_i, x_{-i}) \geq f_i(\|x_i^{+1} - x_i\|)$$

for some "forcing function" $f_i : [0, +\infty[\to [0, +\infty[$ satisfying $f_i(\Delta) \to 0 \Rightarrow \Delta \to 0$.

Applied game theory often offers some agent i to choose merely from a finite set containing, say, n_i (pure) strategies. If so, the mixed extension constrains him to play within the probability simplex

$$X_i = \{x_i \in R_+^{n_i} : \langle x_i, \mathbf{1} \rangle = 1\},$$

thereby getting a payoff $\pi_i(x_i, x_{-i})$ which is linear in own mixed strategy x_i. A most natural direction $d_i = x_i^{+1} - x_i$ for him then comes with $x_i^{+1} \in \arg\max \pi_i(\cdot, x_{-i})$. Except for particular x_{-i} the maximizing choice x_i^{+1} will be *pure* (i.e., equal an extreme point of the simplex X_i). If moreover, at time t he uses stepsize $\rho_{it} = \frac{1}{t+1}$, then we enter the classical realm of so-called *fictitious play* [14], [15], [18]. In that setting x^t is interpreted not as realized play, but rather as a society-wide held belief about strategic intentions at time t. Given this interpretation the iterative steps (24) take on the nature of learning rational expectations. We shall briefly return to this issue in Section 7.

We find the hypothesis (27) somewhat inconvenient. Therefore we now seek to circumvent it.

Definition 9 *(Approximate conditional gradient) We say that individual i uses an ε-approximate conditional gradient rule at x if he opts to move towards a point*

$$x_i^{+1} \in \varepsilon\text{-}\arg\max_{X_i} \langle \nabla_i \pi_i(x), \cdot - x_i \rangle_i$$

with stepsize

$$\rho_i \in \varepsilon\text{-}\arg\max_{[0,1]} \pi_i(x_i + (\cdot)(x_i^{+1} - x_i), x_{-i}).$$

Theorem 7 *(Convergence of approximate conditional gradient play) Suppose*

** the game has a bounded squeezing potential P;*

** each payoff function π_i is continuously differentiable, strictly concave with respect to own decision x_i in some Euclidean space; and that*

** each individual i uses an ε_{it}-approximate conditional gradient rule at time t.*

Then, if $\varepsilon_{it} \to 0, \forall i$, repeated play, in any order, clusters to Nash equilibrium. If moreover, the potential P is exact and jointly concave, then play will accumulate to maximizing points.

Proof. Let $x = \lim_{t \in T} x^t$ be any cluster point of the generated sequence. *We claim* that, possibly after passing to a subsequence, it holds

$$x = \lim_{t \in T} x^{t+1}. \tag{29}$$

Suppose without loss of generality that along T players always act in fixed order, henceforth designated $1, 2, ..., \#I$. Consider now an arbitrary player, say no i. Suppose that for all those preceding him we have already established the desired claim:

$$\lim_{t\in T} x_1^{t+1} = x_1, \quad ... \quad , \lim_{t\in T} x_{i-1}^{t+1} = x_{i-1}.$$

(Evidently, when $i = 1$, this assumption is void.) To produce a contradiction imagine that $\lim_{t\in T} x_i^{t+1} \neq x_i$. Then, introducing

$$\triangle_i^t := \frac{x_i^{t+1} - x_i^t}{\left\| x_i^{t+1} - x_i^t \right\|}, \ \delta_{it} := \left\| x_i^{t+1} - x_i^t \right\|,$$

we may take it that

$$\lim_{t\in T} \triangle_i^t := \triangle_i \neq 0, \text{ and } \delta_{it} \geq \delta_i > 0 \text{ for all } i \in I, t \in T.$$

Fix any $\gamma \in \left]0, 1\right[$. By the convexity of X_i

$$x_i^t + \gamma\delta_i\triangle_i^t \in X_i, \ \forall t \in T.$$

Therefore, letting

$$x^{i,t} := (x_1^{t+1}, ..., x_{i-1}^{t+1}, x_i^t, x_{i+1}^t, ...)$$

be the profile seen by player i at time $t \in T$, the concavity of π_i in x_i implies

$$\varepsilon_{it} + \pi_i(x_i^{t+1}, x_{-i}^{i,t}) \geq \pi_i(x_i^t + \gamma\delta_i\triangle_i^t, x_{-i}^{i,t}) \geq \pi_i(x^{i,t}).$$

Since $P(x_i^{t+1}, x_{-i}^{i,t}) - P(x^{i,t}) \to 0$, use the squeezing (2) and pass to the limit to have

$$\pi_i(x) \geq \pi_i(x_i + \gamma\delta_i\triangle_i, x_{-i}) \geq \pi_i(x).$$

Since $\gamma \in \left]0, 1\right[$ was arbitrary, this contradicts the strict concavity of π_i in x_i. Hence along an appropriate subsequence of T we must also have $x_i^{t+1} \to x_i$. This proves the claim (29). To complete the argument consider an arbitrary agent i, and pick any x_i^{+1} such that $(x_i^{+1}, x_{-i}) \in X$. It suffices now to observe that

$$\begin{aligned}\left\langle \nabla_i\pi_i(x), x_i^{+1} - x_i \right\rangle &= \lim_{t\in T} \left\langle \nabla_i\pi_i(x^t), x_i^{+1} - x_i^t \right\rangle \\ &\leq \lim_{t\in T} \left\{ \left\langle \nabla_i\pi_i(x^t), x_i^{t+1} - x_i^t \right\rangle + \varepsilon_{it} \right\} = 0. \ \Box\end{aligned}$$

7 Playing Infinitely Often

We aim finally at granting agents great flexibility in terms of when - and how large - steps are taken. Such flexibility is offered, in one shot, by the so-called rule of divergent series [7], [8], [10]. A bonus of this approach is a convergence result concerning fictitious play which generalizes recent findings of Monderer and Shapley [14]. Suppose now that

• each marginal payoff function satisfies the Lipschitz condition

$$\|\nabla_i \pi_i(x) - \nabla_i \pi_i(\tilde{x})\| \leq K_i \|x - \tilde{x}\|,$$

stronger than (19), on X, this set being nonempty compact convex and contained in some Euclidean space;

• each player i makes at time t, when seeing the profile $x^{i,t}$, a *payoff-improving choice*

$$x_i^{t+1} = x_i^t + \rho_{it}(\widetilde{x}_i^t - x_i^t) \tag{30}$$

where

$$\widetilde{x}_i^t \in \rho_{it}\text{-}\arg\max_{X_i} \left\langle \nabla_i \pi_i(x^{i,t}), \cdot - x_i^t \right\rangle, \tag{31}$$

and where the stepsizes $\rho_{it} \geq 0$ satisfy

$$\sum_t \rho_{it} = +\infty, \quad \sum_t \rho_{it}^2 < +\infty \tag{32}$$

We start with two useful lemmata:

Lemma 1 *Consider four bounded, nonnegative sequences* $\{\alpha_t\}, \{\beta_t\}, \{\rho_t\}, \{\delta_t\}$ *satisfying*

$$\alpha_t + \beta_{t+1} \geq \beta_t + \rho_t \delta_t,$$

$$\sum \alpha_t < +\infty, \quad \rho_t \to 0, \quad \sum \rho_t = +\infty,$$

and

$$|\delta_{\tau_t} - \delta_t| \to 0 \ \ \text{when} \ \sum_{\tau=t}^{\tau_t - 1} \rho_\tau \to 0,\ \tau_t \geq t, \ \text{and}\ t \to \infty. \tag{33}$$

Then $\{\beta_t\}$ *converges and* $\delta_t \to 0$.

Proof. The convergence of $\{\beta_t\}$ follows easily from $\alpha_t + \beta_{t+1} \geq \beta_t$ and the fact that $\sum \alpha_t < +\infty$.

If $\liminf_{t\to\infty} \delta_t =: \delta > 0$, then $\alpha_t + \beta_{t+1} \geq \beta_t + \rho_t \delta/2$ for all large t, and so the hypotheses $\sum \alpha_t < +\infty$, $\sum \rho_t = +\infty$ imply that $\{\beta_t\}$ must be unbounded, - a contradiction. Thus $\liminf \delta_t = 0$.

We go on to suppose that $\limsup_{t\to\infty} \delta_t =: 2\Delta > 0$. Then there exists a subsequence T such that $\delta_t \geq \Delta$ for all $t \in T$. Define another subsequence $\{\tau_t\}_{t\in T}$ as follows:

$$\tau_t := \min\left\{\tau > t : \delta_\tau \leq \frac{\Delta}{2}\right\}.$$

We have

$$\sum_{\tau=t}^{\tau_t - 1} \alpha_\tau + \beta_{\tau_t} \geq \beta_t + \frac{\Delta}{2} \sum_{\tau=t}^{\tau_t - 1} \rho_\tau.$$

It follows that $\lim_{t\in T}\sum_{\tau=t}^{\tau_t-1}\rho_\tau = 0$, whence via (33) that $|\delta_{\tau_t}-\delta_t| < \frac{\Delta}{2}$ for all large $t \in T$. Thus for such t it holds that

$$\delta_{\tau_t} \geq \delta_t - |\delta_{\tau_t} - \delta_t| > \frac{\Delta}{2},$$

a contradiction. The upshot is that $\Delta = 0$, and the proof is complete. □

Lemma 2 *Let*

$$m_i^t := \max_{X_i} \left\langle \nabla_i \pi_i(x^{i,t}), \cdot - x_i^t \right\rangle. \tag{34}$$

Then for some positive constant C it holds that

$$\left| m_i^{t+1} - m_i^t \right| \leq C \sum_j \{\rho_{jt} + \rho_{jt+1}\}. \tag{35}$$

Proof. Suppose $m_i^{t+1} \leq m_i^t := \langle \nabla_i \pi_i(x^{i,t}), \hat{x}_i^t - x_i^t \rangle$. Then,

$$\begin{aligned}
\left| m_i^{t+1} - m_i^t \right| &= m_i^t - m_i^{t+1} \leq \left\langle \nabla_i \pi_i(x^{i,t}), \hat{x}_i^t - x_i^t \right\rangle - \left\langle \nabla_i \pi_i(x^{i,t+1}), \hat{x}_i^t - x_i^{t+1} \right\rangle \\
&= \left\langle \nabla_i \pi_i(x^{i,t}) - \nabla_i \pi_i(x^{i,t+1}), \hat{x}_i^t - x_i^t \right\rangle - \left\langle \nabla_i \pi_i(x^{i,t+1}), x_i^t - x_i^{t+1} \right\rangle \\
&\leq K_i \left\| x^{i,t} - x^{i,t+1} \right\| \left\| \hat{x}_i^t - x_i^t \right\| + \left\| \nabla_i \pi_i(x^{i,t+1}) \right\| \left\| x_i^t - x_i^{t+1} \right\| \\
&\leq K_i \left\| x^{i,t} - x^{i,t+1} \right\| \left\| \hat{x}_i^t - x_i^t \right\| + \left\| \nabla_i \pi_i(x^{i,t+1}) \right\| \rho_{it} \left\| \tilde{x}_i^t - x_i^t \right\| \\
&\leq \left\{ K_i \sum_j \{\rho_{jt} + \rho_{jt+1}\} \max_j diam X_j + \left\| \nabla_i \pi_i(x^{i,t+1}) \right\| \rho_{it} \right\} diam X_i
\end{aligned}$$

and now (35) follows. The case $m_i^{t+1} \geq m_i^t$ is handled in the same manner. □

Theorem 8 *(Convergence with decreasing, divergent stepsizes) Suppose the game has a bounded potential satisfying the sandwich property (3). Also suppose directions are close to best possible in the sense (31). Then, under (32), repeated play of the payoff improving sort (30), performed in any order, will cluster to stationary points.*

Proof. The same argument as after (21) gives

$$\begin{aligned}
w_i \left\{ P(x^{t+1}) - P(x^t) \right\} &\geq \pi_i(x_i^{t+1}, x_{-i}^{i,t}) - \pi_i(x^{i,t}) \\
&\geq \rho_{it} \left\langle \nabla_i \pi_i(x^{i,t}), \tilde{x}_i^t - x_i^t \right\rangle - \rho_{it}^2 \frac{K_i}{2} \left\| \tilde{x}_i^t - x_i^t \right\|^2 \\
&\geq \rho_{it} m_i^t - \rho_{it}^2 \frac{K_i}{2} \left\| \tilde{x}_i^t - x_i^t \right\|^2 - \rho_{it}^2
\end{aligned}$$

with

$$m_i^t := \max \left\langle \nabla_i \pi_i(x^{i,t}), \cdot - x_i^t \right\rangle = \max \left\langle \nabla_i P(x^{i,t}), \cdot - x_i^t \right\rangle.$$

Now invoke Lemma 1 with

$$\alpha_t = \rho_{it}^2 \frac{K_i}{2} \left\| \tilde{x}_i^t - x_i^t \right\|^2 + \rho_{it}^2, \quad \beta_t = w_i P(x^t), \quad \rho_t = \rho_{it}, \quad \text{and} \quad \delta_t = m_i^t.$$

Lemma 2 furnishes (35) which in turn implies (33). Thus, $m_i^t \to 0, \forall i$, and the proof is complete. □

Corollary 3 *(Convergence of fictitious play). Suppose a finite game has a potential satisfying the sandwich property (3). Then fictitious play, in any order, clusters to Nash equilibrium.*

Proof. All assumptions of the preceding theorem are fulfilled: First, individual payoffs are linear in own variable. Second, direction finding reduces to best response. Third, the stepsizes $\rho_{it} = \frac{1}{t+1}$ satisfy (32). □

It follows from this last result that many games, known to produce divergent fictitious play, cannot be potential. For more on this topic see [11], [15], [17].

8 Concluding Remarks

We have always let $X = \Pi_{i\in I} X_i$ be a product of individual decision spaces. Often there are coupling constraints though, affecting admissible strategy profiles x. Examples include restrictions derived from concerns with efficiency, equity, social security, financial viability, etc. Then, X will be strictly contained in the hyperbox $\Pi_{i\in I} X_i$, but is presumably still closed convex. If so, we may still use the methods described above as long as each agent i takes care that any adjustment rule $\mathfrak{A}_i$ of his satisfies

$$\mathfrak{A}_i(x) \subseteq \left\{x_i^{+1} : (x_i^{+1}, x_{-i}) \in X\right\}.$$

For example, one could replace gradient projection (18) by

$$\mathfrak{A}_i(x) := [x_i + \rho_i \nabla_i \pi_i(x)]^+_{x_{-i}},$$

where $[\cdot]^+_{x_{-i}}$ now denotes orthogonal projection onto the set $\left\{x_i^{+1} : (x_i^{+1}, x_{-i}) \in X\right\}$. Likewise, given adjustment costs (11), the rule (16) should come in the form

$$\mathfrak{A}_i(x) := \left\{x_i^{+1} : (x_i^{+1}, x_{-i}) \in X \text{ and } \pi_i(x_i^{+1}, x_{-i}) - a_i(x_i^{+1}, x_i) \text{ is maximal}\right\}.$$

It must be emphasized though that, given coupling constraints, other ideas, based on duality and Lagrange multipliers, naturally enter the field. Also, for any method, whether intended as a story about repeated play - or preplay, or simply offered as an algorithm - one would like to know the rate of convergence.

It also deserves mention that uncertainty can be accommodated [9]. However, to discuss such issues would take us outside reasonable confines of this paper.

References

[1] L. Armijo, Minimization of functions having continuous partial derivatives, *Pacific Journal of Mathematics* 16, 1-3 (1966).

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation,* Prentice Hall, Englewoods Cliffs, New Jersey (1989).

[3] D. P. Bertsekas, *Nonlinear Programming,* Athena Scientific, Belmont, Mass. (1995).

[4] E. Cavazzuti, Convergence of equilibria in the theory of games, *Lecture Notes in Mathematics* 1190, Springer-Verlag, Berlin 95-103 (1986).

[5] E. Cavazzuti and N. Pacchiarotti, Convergence of Nash equilibria, *Boll. Un. Mat. Ital.* B5, 247-266 (1986).

[6] G. Dal Maso, *An Introduction to* $\Gamma-$*Convergence,* Birkhaüser, Basel (1993).

[7] Y. M. Ermoliev, Methods for solving nonlinear extremal problems, *Cybernetics* 2, 1-14 (1966).

[8] Y. M. Ermoliev and S. P. Uryasiev, Nash equilibrium in n-person games, *Kibernetika* 3,85-88 (1982).

[9] Y. M. Ermoliev and R. J-B. Wets (eds.), *Numerical Techniques for Stochastic Optimization,* Computational Mathematics 10, Springer-Verlag, Berlin (1988).

[10] S. D. Flåm, Approaches to economic equilibrium, *Journal of Economic Dynamics and Control* 20, 1505-1522 (1996).

[11] D. P. Foster and H. Peyton Young, On the nonconvergence of fictitious play in coordination games, *Working paper,* IIASA, Laxenburg 33 (1996).

[12] D. Fudenberg and D. K. Levine, *Theory of Learning in Games,* Book draft (1996).

[13] D. Monderer and L. S. Shapley, Potential games, *Games and Economic Behavior* 14, 124-143 (1996).

[14] D. Monderer and L. S. Shapley, Fictitious play property for games with identical interest, *Journal of Economic Theory* 68, 258-265 (1996).

[15] D. Monderer and A. Sela, A 2×2 game without the fictitious play property, *Games and Economic Behavior,* 14, 144-148 (1996).

[16] H. Peyton Young, *Learning and Evolution in Games,* Book draft (1996).

[17] M. Posch, Cycling with a generalized urn scheme and a learning algorithm for 2×2 games, *Working paper,* IIASA, Laxenburg 76 (1994).

[18] J. Robinson, An iterative method of solving a game, *Ann. Math.* 54, 296-301 (1951).

[19] R. T. Rockafellar, *Convex Analysis,* Princeton University Press, New Jersey (1970).

[20] R. W. Rosenthal, A class of games possessing pure-strategy Nash equilibria, *Int. J. Game Theory* 2, 65-67 (1973).

[21] T. Sjöström and M. L. Weitzman, Competition and the evolution of efficiency, *Journal of Economic Behavior & Organization* 30, 25-43 (1996).

[22] M. Teboulle, Entropic proximal mappings with applications to nonlinear programming, *Mathematics of Operations Research* 17, 670-690 (1992).

[23] F. Vega-Redondo, Competition and culture in an evolutionary process of equilibrium selection: A simple example, *Journal of Games and Economic Behavior* 5, 618-631 (1993).

[24] W. I. Zangwill, *Nonlinear Programming, A unified approach,* Prentice Hall, Englewoods Cliffs, New Jersey (1969).