# Stochastic Learning of Equilibria in Games: The Ordinary Differential Equation Method

Olivier Bournez[1] and Johanne Cohen[2]

[1] Ecole Polytechnique,
Laboratoire LIX, F-91128 Palaiseau Cedex,
Olivier.Bournez@lix.polytechnique.fr
[2] Centre National de la Recherche Scientifique,
Laboratoire PRiSM Université de Versailles 45, avenue des Etats-Unis, F-78000 Versailles,
Johanne.Cohen@prism.uvsq.fr

**Abstract.** Our purpose is to discuss stochastic algorithms to learn equilibria in games, and their time of convergence. To do so, we consider a general class of stochastic algorithms that converge weakly (in the sense of weak convergence for stochastic processes) towards solutions of particular ordinary differential equations, corresponding to their mean-field approximations. Tuning parameters in these algorithms provides several dynamics having limit points related to Nash equilibria, and hence provide means to compute equilibria in a distributed fashion in games. We relate the time of convergence of stochastic dynamics to the time of convergence of their corresponding ordinary differential equation. This gives lower and upper bounds on the time needed to learn equilibria in games through such stochastic dynamics.

## 1 Introduction

In many situations, some actors can own profit by using some resources according to their own interest to the detriment of a better social behavior. Peer-to-Peer applications, with phenomena such as free-riding, provide an example of problematic behaviors in large-scale distributed systems. Routing in Internet network, where some economic interests may provide incentives to alter global performances, is another example.

In all these contexts, agents often adapt their strategies based on their local knowledge of the system, by small adjustments in order to improve their own profit. The impact of each individual player on the network is small. However, as the number of actors is large, a global evolution of the system may happen.

We are interested in understanding when the system can converge towards rational situations, i.e. Nash equilibria in the sense of game theory. This is natural to expect dynamics of adjustments to be stochastic, and fully distributed, since agents are often involved in games where a local, deterministic description of the whole system is not possible, and since decisions are often attempts and errors guided.

Several such dynamics of adjustments have been considered recently in the algorithmic game theory literature. Up to our knowledge, this has been done mainly for deterministic dynamics and mainly for best-response based dynamics in algorithmic game theory: Computing a best response requires a global description of the system. Stochastic variations, avoiding a global description, have been considered. However, considered dynamics are somehow rather ad-hoc, in order to get efficient convergence time bounds, and still mainly best-response based. We want to consider here more general dynamics, and discuss when one may expect convergence, with time bounds on their convergence.

Our settings is the following: Let $[n] = \{1, \dots, n\}$ be the set of players. Every player $i$ has a set $\mathcal{S}_i$ of *pure strategies*. Let $m_i$ be the cardinal of $\mathcal{S}_i$. A *mixed strategy* $q_i = (q_{i,1}, q_{i,2}, \dots, q_{i,m_1})$ corresponds to a probability distribution over pure strategies: pure strategy $\ell$ is chosen with probability $q_{i,\ell} \in [0, 1]$, with $\sum_{\ell=1}^{m_i} q_{i,\ell} = 1$. Let $K_i$ be the simplex of mixed strategies for player $i$. Any pure strategy $\ell$ can be considered as mixed strategy $e_\ell$, where vector $e_\ell$ denotes the unit probability vector with $\ell^{th}$ component unity, hence as a corner of $K_i$. Let $K = \prod_{i=1}^{n} K_i$ be the space of all mixed strategies. A *strategy profile* $Q = (q_1, ..., q_n) \in K$ specifies the (mixed or pure) strategies of all players: $q_i$ corresponds to the mixed strategy played by player $i$. Following classical convention, we write abusively $Q = (q_i, Q_{-i})$, where $Q_{-i}$ denotes the vector of the strategies played by all other players.

Games with random payoffs are allowed: we only assume that whenever the strategy profile $Q \in K$ is known, each player $i$ gets a random *cost* of well-defined expected value $c_i(Q)$. In particular, the expected cost for player $i$ for playing pure strategy $e_\ell$ is denoted by $c_i(e_\ell, Q_{-i})$. Strategy $x$ of player $i$ is said to be a *best reply*, or *best-response* to $Q_{-i}$ if $c_i(x, Q_{-i}) \leq c_i(q_i, Q_{-i})$ for all $q_i \in K_i$. The (compact, convex, non-empty) set of all best replies to the strategy profile $Q$ of player $i$ is denoted by $BR_i(Q)$.

We want basically to consider learning algorithms of the following form, over the most possible general games, where $b$ is a parameter, intended to be positive but close to 0. Functions $\sigma_i$, can be considered to be identity in a first step, but can be actually any function with positive values, with $\sum_\ell \sigma_{i,\ell}(q_i) = 1$.

– Initially, $q_i(0) \in K_i$ can be any vector of probability, for all $i$.
– At each round $t$, for each player $i$:
  - selects a strategy $s_i(t) \in \mathcal{S}_i$ according to distribution $\sigma_i(q_i(t))$: player $i$ selects strategy $\ell \in \mathcal{S}_i$ with probability $\sigma_{i,\ell}(q_{i,\ell}(t))$.
  - This leads to a (random) cost $r_i(t)$ for player $i$.
  - Updates $q_i(t)$ as follows:

$$q_i(t + 1) = q_i(t) + bF_i^b(r_i(t), s_i(t), q_i(t)). \qquad (1)$$

Let $Q(t) = (q_1(t), ..., q_n(t)) \in K$ denote the state of all players at instant $t$. Our interest is in the asymptotic behavior of $Q(t)$, and its possible convergence to Nash equilibria. When $Y$ is some random variable, we write $\mathbb{E}_\sigma[\ Y\ |Q\ ]$ for the expectation of $Y$ when players play according to distribution $\sigma(Q)$: that is to say, when player $j$ chooses strategy $\ell \in \mathcal{S}_j$ with probability $\sigma_{j,\ell}(q_{j,\ell})$.

Functions $F_i^b(r_i(t), s_i(t), q_i(t)))$ can be as generic as possible, assuming that the $q_i(t)$ always stay validity probability vectors: that is to say, $q_{i,\ell}(t) \in [0, 1]$ and $\sum_\ell q_{i,\ell}(t) = 1$ is preserved. In particular, functions $F_i^b(r_i(t), s_i(t), q_i(t))$ can be random valued[1]. We only assume that their expectation $\mathbb{E}_\sigma[\ F_i^b(r_i(t), s_i(t), q_i(t))\ |Q(t)\ ]$ is always defined, and that $G_i(Q) = \lim_{b\to 0} \mathbb{E}_\sigma[\ F_i^b(r_i(t), s_i(t), q_i(t))\ |Q\ ]$ exists and is a continuous function of $Q$.

Algorithms of this form correspond indeed to fully distributed algorithms. Decisions made by players are completely decentralized: At each time step, player $i$ only needs $r_i$ and $q_i$, that is to say respectively her current cost and her current mixed strategy, to update her own strategy $q_i$.

[1] In this case, formally, their may depend on $Q$, through some “local” experiments.

## 2 Summary of Results

In the general case (Theorem 1), any stochastic algorithm in the considered class converges weakly (in the sense of weak convergence for probabilistic processes) towards solutions of initial value problem (ordinary differential equation (ODE))

$$\begin{cases} \frac{dX}{dt} & = G(X) \\ X(0) & = Q(0) \end{cases} \tag{2}$$

where $G(Q) = (G_1(Q), \cdots, G_n(Q))$. In other words, any stochastic algorithm is weakly-convergent to its mean-field limit approximation given by (2).

This gives ways to obtain dynamics whose limit points are related to Nash equilibria.

As a first example (Section 5.1), take functions $\sigma_i$ to be identity, and $F_i^b(r_i(t), s_i(t), q_i(t)))$ defined as follows:

$$F_i^b(r_i(t), s_i(t), q_i(t)) = \gamma(r_i(t))(e_{s_i(t)} - q_i(t)) \tag{3}$$

where $\gamma : \mathbb{R} \to [0,1]$ is some affine decreasing function with value in $[0,1]$. Recall that $e_{s_i(t)}$ is the unit vector of dimension $m_i$ with component number $s_i(t)$ unity.

It follows that the stochastic algorithms behaves like ordinary differential equation

$$\frac{dq_{i,\ell}}{dt} = -q_{i,\ell}(c_i(e_\ell, Q_{-i}) - c_i(q_i, Q_{-i})), \tag{4}$$

that is to say a (multi-population) replicator dynamic, whose limit points (Proposition 1) are well-known to be related to Nash equilibria (through so-called Folk's theorems of evolutionary game theory [13]).

As a second example (Section 5.2), take $\sigma_i$ given componentwise by

$$\sigma_{i\ell}(q_i) = \frac{\exp(q_{i,\ell}/\kappa)}{\sum_j \exp(q_{i,j}/\kappa)}, \tag{5}$$

that is to say is the *logit dynamics*, where $\kappa$ is some positive constant. When $\kappa$ goes to 0 this converges to the best response dynamics, whereas when $\kappa$ goes to infinity, this acts as some uniform choice [13].

Consider $F_i^b(r_i(t), s_i(t), q_i(t))$ given by

$$F_i^b(r_i(t), s_i(t), q_i(t)) = (\gamma(r_i(t)) - q_{i,s_i(t)})e_{s_i(t)} \tag{6}$$

Limit points (Proposition 2) of the associated ordinary differential equation correspond to Nash equilibria of a game whose payoffs are perturbed by an additive term that goes to 0 when $\kappa$ goes to 0.

As a third example, take functions $\sigma_i$ to be identity, and $F_i^b(r_i(t), s_i(t), q_i(t)))$ defined as follows: (i) choose some other pure strategy $e_j(t)$ uniformly at random: strategy $j$ is chosen with probability $\frac{1}{m_i}$; (ii) plays according to this strategy $e_j(t)$. This leads to a (random) cost $r_j(t)$ for player $i$; (iii) Consider then

$$F_i^b(r_i(t), s_i(t), q_i(t))) = \nu(r_i(t) - r_j(t))(e_j - q_i(t)),$$

where $\nu$ is some increasing function which is 0 for negative argument, and 1 for argument greater than $\epsilon > 0$. This leads to a dynamic close to the one considered in [2]. As long as the current state $Q(t)$ is not in a $\epsilon$-Nash equilibrium, this moves to a better response. Stationary points must correspond to $\epsilon$-Nash equilibria.

In all these examples, we may expect the stochastic algorithms to behave like associated mean-field approximation (2). As all of them have stable stationary points corresponding to Nash equilibria, if for the considered game this dynamic is convergent, then we can expect the stochastic algorithms to converge towards Nash equilibria.

Notice, that there is no reason that convergence of mean-field approximation holds for generic games, but if it holds, then its stable limit points will be Nash equilibria (or $\epsilon$-Nash equilibria for the third).

Similar facts have been established several times in game theory literature for specific cases of stochastic algorithms [22, 16, 9]. Proofs, often rely on the theory of approximation of stochastic sequence by weak-convergence methods, and do not provide time bounds.

We prove (Theorem 3) that is indeed possible to go further and that it is indeed possible to talk about the time of convergence in the general case. In other words, the time of convergence of the stochastic algorithm can be closely related to the time of convergence of the associated ordinary differential equation. This provides upper-bounds and lower-bounds on the time before convergence of considered stochastic algorithms.

Notice that this is established through some bounds of very general interest (Theorem 2).

We do a generic study of symmetric $2 \times 2$ games in Section 8, to show that one may expect a convergence in a time polynomial in $\frac{1}{\epsilon}$ for such games.

## 3 Related work.

In algorithmic game theory, much attention have been put on particular ordinal potential games, and on (exact) potential games. Following [20], a game is an *ordinal potential game* if there exists some function whose sign of variations reflects the sign of variation of utility of any player doing a pure strategy move. A game is an *exact potential game* if there exists some function whose variations reflect exactly the variation of utility of any player doing a pure strategy move. In particular, an (exact) potential game is an ordinal potential game.

Load balancing games, introduced in [15] are restricted instances of congestion games. Congestion games, introduced in [23], are known to be particular exact potential games [23]. Actually, it is known that a game is an exact potential game iff it is isomorphic to a congestion game [20]. Ordinal potential games include task allocation games introduced in [4].

An ordinal potential game always have a pure Nash equilibrium: since an ordinal potential function, that can take only a finite number of values, is strictly decreasing in any sequence of pure strategy strict best response moves, such a sequence must be finite and must lead to a Nash equilibrium [23]. This proof of existence of pure Nash equilibria can be turned into a dynamic: players play in turn, and move to resources with a lower cost.

For (exact) potential games, bounds on the time of convergence of best-response dynamics have been investigated following this idea in [6]. Since players play in turns, this is often called the *Elementary Stepwise System*. Other results of convergence in this model, have been investigated in [10, 18, 21], but all require some global knowledge of the system in order to determine what next move to choose.

For load balancing games, a Stochastic version of best-response dynamics has been investigated in [2]. The expected time of convergence to an $\epsilon$-Nash equilibrium is in $\mathcal{O}(nmW^3\epsilon^2)$ where $W$ denotes the maximum weight of any task.

For congestion games, the problem of finding pure Nash equilibria is known to be PLS-complete [14]. Efficient convergence of particular best-response dynamics to approximate Nash equilibria in symmetric congestion games has been investigated in [3], in the particular case where each resource cost function satisfies a *bounded jump assumption*. In this context, the convergence to an $\epsilon$-Nash equilibrium occurs within a number of steps that is polynomial in the number of players.

The stochastic dynamic (3) has been partially investigated in [22] for general games and for potential games: It has been proved to be weakly convergent to solutions of a multipopulation replicator equation. Notice that compared to [22], we allow perturbed dynamics, and we discuss time of convergence.

Replicator equations have been deeply studied in evolutionary game theory [13, 26]. Evolutionary game theory has been applied to routing problems in the Wardrop traffic model in [8, 7]. Evolutionary game theory doesn't restrict to above discussed dynamics, but considers a whole family of dynamics that satisfy Folk's theorems in the spirit of Proposition 1.

Bounds on the rate of convergence of fictious play dynamics have been established in [11]. Fictious play has been reproved to be convergent for zero-sum games using numerical analysis methods, or more generally stochastic approximation theory: fictious play can be proved to be an Euler discretization of a certain continuous time process [13].

The proof of Theorem 1 relies on weak-convergence methods for probabilistic processes. Some studies of time of convergence have been made in this context (see e.g. monograph [17]) but one must understand that what is understood as a time of convergence from a probabilistic point of view generally differs from the computer science point of view: time of convergence often means in a probabilistic settings central-limit like theorems for the limit processes, whereas this kind of results, still mainly often obtained by weak-convergence methods, do not help to talk about the distance between the process and its mean-field approximation.

The approximation result obtained in Theorem 2 is obtained from ideas of constructions from [1]. The bounds that we obtain are tuned up to our context, and somehow doubly perturbed (random+deterministic perturbation) whereas [1] considers only random perturbations, and state some very generic approximation results, stating that error is function of the time variable, without really taking care on the exact dependence.

## 4 Weak-Convergence Results

Recall that we are interested in discussing the evolution of $Q(t)$, where $Q(t) = (q_1(t), ..., q_n(t)) \in K$ denotes the state of the player team at instant $t$ in the stochastic algorithm. Clearly, $Q(t)$ is an homogeneous Markov chain. Define $\Delta Q(t)$ as $\Delta Q(t) = Q(t+1) - Q(t)$, and $\Delta q_i(t)$ as $q_i(t+1) - q_i(t)$. We can write

$$\mathbb{E}[\ \Delta q_i(t)\ |Q(t)\ ] = b\mathbb{E}_\sigma[\ F_i^b(r_i(t), s_i(t), q_i(t))\ |Q(t)\ ], \tag{7}$$

with $G_i(Q) = \lim_{b\to 0} \mathbb{E}_\sigma[\ F_i^b(r_i(t), s_i(t), q_i(t))\ |Q(t)\ ]$ assumed to be continuous.

Let us first state informally, why we expect $Q(t)$ to converge towards solutions of ordinary differential equation (2): Assume we replace $\mathbb{E}[\ \Delta q_i(t)\ |Q(t)\ ]$ by $\Delta q_i(t)$ in this last equation. Through the change of variable $t \leftarrow tb$, this would become $q_i(t+b) - q_i(t) = b\mathbb{E}_\sigma[\ F_i^b(r_i(t), s_i(t), q_i(t))\ |Q(t)\ ]$. Approximating $q_i(t+b) - q_i(t)$ by $b\frac{dq_i}{dt}(t)$ for small $b$, we may expect the system to behave like ordinary differential equation (ODE) $\frac{dq_i}{dt} = G_i(Q)$, when $b$ is close to 0.

This can be formalized as follows: Consider the piecewise-linear interpolation $Q^b(.)$ of $Q(t)$ defined by

$$Q^b(t) = Q(\lfloor t/b \rfloor) + (t/b - \lfloor t/b \rfloor)(Q(\lfloor t/b + 1 \rfloor) - Q(\lfloor t/b \rfloor)). \tag{8}$$

Function $Q^b(.)$ belongs to the space of all functions from $\mathbb{R}$ into $K$ which are right continuous and have left hand limits (*cad-lag functions*). Now consider the sequence $\{Q^b(.) : b > 0\}$. We are interested in the limit $Q(.)$ of this sequence when $b \to 0$. Recall that a family of random variable $(Y_t)_{t \in \mathbb{R}}$ weakly converges to a random variable $Y$, if $E[h(X_t)]$ converges to $E[h(Y)]$ for each bounded and continuous function $h$.

**Theorem 1.** *The sequence of interpolated processes $\{Q^b(.)\}$ converges weakly, when $b \to 0$, to $Q(.)$, which is the (unique deterministic) solution of initial value problem* (2)*.*

This can be proved using weak-convergence methods. Formally, this follows from [25, theorem 11.2.3], once this is understood that all the hypotheses of this general result hold[2].

# 5 Examples of Dynamics to Learn Equilibria

## 5.1 A Replicator Dynamics

Suppose we consider functions $\sigma_i$ as the identity, and a dynamics for $F_i^b(r_i(t), s_i(t), q_i(t))$ as in (3). If one prefers, componentwise:

$$\Delta q_{i,\ell}(t) = q_{i,\ell}(t+1) - q_{i,\ell}(t) = \begin{cases} -b\gamma(r_i(t))q_{i,\ell}(t) & \text{if } s_i(t) \neq \ell \\ -b\gamma(r_i(t))q_{i,\ell}(t) + b(\gamma(r_i(t))) & \text{if } s_i(t) = \ell. \end{cases}$$

We then have

$$\begin{aligned} G_{i,\ell}(Q) &= \lim_{b \to 0} \tfrac{1}{b} \mathbb{E}[\ \Delta q_{i,\ell}(t)\ |Q(t)\ ] \\ &= \lim_{b \to 0} \tfrac{1}{b} (\textstyle\sum_j q_{i,j}(t) \mathbb{E}[\ \Delta q_{i,\ell}(t)\ |Q(t), s_i(t) = j\ ]) \\ &= +\textstyle\sum_j q_{i,j}(t) q_{i,\ell}(t) \mathbb{E}[\ \gamma(r_i(t))\ |Q(t), s_i(t) = \ell\ ]) \\ &\quad -\textstyle\sum_j q_{i,j}(t) q_{i,\ell}(t) \mathbb{E}[\ \gamma(r_i(t))\ |Q(t), s_i(t) = j\ ]) \\ &= q_{i,\ell}(\mathbb{E}[\ \gamma(r_i(t))\ |Q(t), s_i(t) = \ell\ ] - \mathbb{E}[\ \gamma(r_i(t))\ |Q(t)\ ]). \end{aligned}$$

In other words, Equation (2) leads by Theorem 1 to (some rescaling introduced by affine decreasing function $\gamma$ of) Dynamics (4). This equation, called the (multipopulation) replicator dynamics, is well-known to have its limit points related to Nash equilibria (through so-called Folk's theorems of evolutionary game theory [13]). More precisely, we have:

**Proposition 1.** *The following are true for the solutions of Equation* (4)*: (i) All Nash equilibria are stationary points. (ii) All strict Nash equilibria are asymptotically stable. (iii) All stable stationary points are Nash equilibria.*

Actually, all corners of simplex $K$ are stationary points, as well as, from the form of (4), more generally any state $Q$ in which all strategies in its support perform equally well. Such a state $Q$ is not a Nash equilibrium as soon as there is an not used strategy (i.e. outside of the support) that performs better.

[2] See Appendix A.

Unstable limit stationary points may exist for the mean-field approximation (4): Consider for example a dynamics that leave on some face of $K$ where some well-performing strategy is never used. To avoid "bad" (non-Nash equilibrium, hence unstable) stationary points, following the idea of penalty functions for interior point methods, one can use as in Appendix A.3 of [24] some patches on the dynamics that would guarantee Non-complacency. *Non-Complacency (NC)* is the following property: $G(Q) = 0$ implies that $Q$ is a Nash equilibrium of (4) (i.e. stationary implies Nash).

For general games, we get that the limit for $b \to 0$ of the dynamics of stochastic algorithms is some ordinary differential equation whose stable limit points, when $t \to \infty$, if there exist, can only be Nash equilibria. Hence, if there is convergence of the ordinary differential equation, then one expects the previous stochastic algorithms to learn equilibria.

### 5.2 A Smoothed Best Response Dynamics

Consider as a second example the case where $\sigma_i$ is given componentwise by Equation (5), and a dynamics of the form (6).

We have in this case

$$\begin{aligned} G_{i,\ell}(Q) &= \lim_{b \to 0} \tfrac{1}{b} \mathbb{E}_\sigma[\ \Delta q_{i,\ell}(t)\ |Q(t)\ ] \\ &= \lim_{b \to 0} \tfrac{1}{b} (\textstyle\sum_j \sigma_{i,j}(q_{i,j}(t)) \mathbb{E}_\sigma[\ \Delta q_{i,\ell}(t)\ |Q(t), s_i(t) = j\ ]) \\ &= \sigma_{i,\ell}(q_{i,\ell}(t))(\mathbb{E}_\sigma[\ \gamma(r_i(t))\ |Q(t), s_i(t) = \ell\ ] - q_{i,\ell}). \end{aligned}$$

In other words, Equation (2) leads by Theorem 1 to

$$\frac{dq_{i,\ell}}{dt} = \sigma_{i,\ell}(q_{i,\ell}(t))(\mathbb{E}_\sigma[\ \gamma(r_i(t))\ |Q(t), s_i(t) = \ell\ ] - q_{i,\ell}). \tag{9}$$

Assume to simplify discussion that $\gamma(x) = -x$ (otherwise, some rescaling+isomorphism might be necessary). Using the fact that logit dynamic $\sigma_i$ correspond to the unique maximizer of strictly concave function $q_i \to \sum_\ell z_\ell q_{i,\ell} - \kappa \sum_\ell z_\ell \log z_\ell$ over the interior of $K$ (see [13]), we get:

**Proposition 2.** *Stationary points of Dynamics* (9) *corresponds to Nash equilibria of the game whose payoffs are given by* $v_i(Q) = c_i(Q) - \kappa \sum_\ell q_{i,\ell} \log q_{i,\ell}$.

In other words, limit points of the associated ordinary differential equation correspond to Nash equilibria of a game whose payoffs are perturbed by a (entropy) term that goes to 0 when $\kappa$ goes to 0. Once again, for general games, we get that the limit for $b \to 0$ of the dynamics of stochastic algorithms is some ordinary differential equation whose stable limit points, when $t \to \infty$, if there exist, can only be Nash equilibria. Hence, if there is convergence of the ordinary differential equation, then one expects the previous stochastic algorithms to learn equilibria.

## 6 Finite Horizon Time Bounds on Stochastic Algorithms

The default of weak-convergence results in the spirit of Theorem 1 is that they do not provide any way to talk about the speed of convergence: This is not possible to tell, given $\epsilon > 0$, what is the time $T$ required to be at distance $\epsilon$ from the ordinary differential equation, nor what should be the discretization step $b$.

We now prove that this is possible to bound the error committed by approximating the stochastic algorithm by its mean-field approximation (2).

First observe that we can rewrite Dynamics (1) of the algorithm as

$$Q(t+1) = Q(t) + bG(Q) + bD(Q) + bR(Q) \tag{10}$$

where

$$D(Q) = \mathbb{E}_\sigma[\ F_i^b(r_i(t), s_i(t), q_i(t))\ |Q(t)\ ] - G_i(Q)$$

can be seen as a deterministic perturbation and

$$R(Q) = F_i^b(r_i(t), s_i(t), q_i(t)) - \mathbb{E}_\sigma[\ F_i^b(r_i(t), s_i(t), q_i(t))\ |Q(t)\ ]$$

as a random perturbation. By hypotheses, $D(Q)$ is in $\mathcal{O}(b)$ and $\mathbb{E}[\ R(Q)\ |Q\ ] = 0$.

In absence of these perturbations, Dynamic (10) would behave like Euler's approximation scheme $q_i(t+1) = q_i(t) + bG_i(Q)$ for solving ordinary differential equation (2). The idea is hence to use generalizations of bounds on the error committed by Euler's method for solving differential equations, taking into accounts perturbations.

This gives the following technical result, of very general interest, whose proof is inspired by constructions from [1].

**Theorem 2.** *Assume a sequence $Q(t) = (q_1(t), \cdots, q_n(t)$, $t \in \mathbb{N}$, leaving in a compact $K$, satisfies a relation of the form* (10).

*Assume that (i) $G$ is a function of class $\mathcal{C}^1$, (ii) $D(.)$ is a deterministic function, of order $\mathcal{O}(b)$: it takes values in $[-Nb, Nb]$ for some constant $N$. (iii) $R(.)$ is some random variable with $\mathbb{E}[\ R(Q)\ |Q\ ] = 0$ for all $Q$, and that random variable $R(Q(k))$ is measurable in the $\sigma$-field generated by $(Q(t))_{0 \leq t \leq k}$, for all $k$.*

*Then, let $X(t)$ be the solution of initial value problem* (2).

*With high probability,*

$$||Q(k) - X(kb)|| \leq be^{\Lambda kb}(A + bB),$$

*for all integer $k$, for some constants $A$, $B$ and $\Lambda$ that do not depend on $b$ and $k$ (and actually only of function $G$).*

*This probability can be made greater than $\mu$ for all $\mu < 1$.*

*Proof.* Since $X(t)$ is the solution of initial value problem (2), we have $X(t) - Q(0) = \int_0^t G(X(t))dt$.

Denote by $\overline{D}$, $\overline{R}$, $\overline{Q}$, $\overline{X}$ the piecewise constant interpolation of $D$, $R$ and $Q$ respectively: they are defined on $\mathbb{R}^{\geq 0}$ by $\overline{D}(t) = D(k)$, $\overline{R}(t) = R(k)$, $\overline{Q}(t) = Q(k)$, $\overline{X}(t) = X(k)$ for any $t \in [kb, (k+1)b)$.

Recall the piecewise-linear interpolation $Q^b(.)$ of $Q$ introduced in Equation (8). Using these functions, Dynamics (10) can be rewritten as

$$Q^b(t) - Q(0) = \int_0^t [G(\overline{Q}(t)) + \overline{D}(t) + \overline{R}(t)]dt \tag{11}$$

If we introduce $\epsilon(t) = Q^b(t) - X(t)$, we then have $\epsilon(t) - \epsilon(0) = \epsilon(t) = T_1 + T_2 + T_3 + T_4$ where

$$T_1 = \int_0^t [G(Q^b(t)) - G(X(t))]dt, \quad T_2 = \int_0^t [G(\overline{Q}(t)) - G(Q^b(t))]dt,$$

$$T_3 = \int_0^t \overline{D}(t)dt, \quad T_4 = \int_0^t \overline{R}(t)dt.$$

$G$ being $\mathcal{C}^1$ on compact $K$, it is $\Lambda$-Lipschtiz for some constant $\Lambda$. The norm of $T_1$ is hence bounded by $\Lambda \int_0^t ||\epsilon(t)||dt$, and the norm of $T_2$ is bounded by $\Lambda \int_0^t ||\overline{Q}(t) - Q^b(t)dt||$.

Now, from Equation (11),

$$||Q^b(t) - \overline{Q}(t)|| = || \int_{b\lfloor t/b \rfloor}^{t} [G(\overline{Q}(t)) + \overline{D}(t) + \overline{R}(t)]dt|| \leq bP + b^2N + bM,$$

where $P$ is a bound of $G$, that exists since $G$ is continuous on compact $K$, and $M$ is a bound on $R$ on compact $K$: recall that a compact is necessarily bounded, and hence that $R$ must take value in $[-M, M]$ for some integer $M$. Hence, the norm of $T_2$ is bounded by $\Lambda tb[P + bN + M]$.

The norm of $T_3$ is less than

$$\int_0^{b\lceil t/b \rceil} ||\overline{D}(t)||dt = \sum_{k=0}^{\lceil t/b \rceil} b||D(k)|| \leq b^2N[\lceil t/b \rceil + 1] \leq bN(t + 2b).$$

Consider sequence $Z_n = \sum_{k=0}^{n-1} R(k)$, with $Z_0 = 0$. $Z_n$ is a martingale[3]. From the fact that $R(k)$ leaves in $[-M, M]$, it has bounded variations: We can then apply Azuma-Hoeffding's Inequality (Lemma 2 in Appendix) to get, that for all $t \geq 0$, and all $\lambda > 0$,

$$Pr(||Z_t - Z_0|| \geq \lambda) = Pr(|| \sum_{k=0}^{t-1} R(k)|| \geq \lambda) \leq 2e^{-\lambda^2/(2tM^2)}.$$

Using some union bounds,

$$Pr( \sup_{0 \leq k \leq n} || \sum_{k=0}^{n} R(k)|| > \lambda) \leq Pr( \bigcup_{0 \leq k \leq n} || \sum_{k=0}^{n} R(k)|| > \lambda) \leq \sum_{k=0}^{n} Pr(|| \sum_{j=0}^{k} R(k)|| > \lambda),$$

which is less than

$$\sum_{k=0}^{n} 2e^{-\lambda^2/(2(k+1)M^2)} \leq 2(n+1)e^{-\lambda^2/(2M^2)}.$$

So, with high probability, that is probability greater than $1 - 2(n+1)e^{-\lambda^2/(2M^2)} \geq 1 - \mathcal{O}(\frac{t}{b}e^{-\lambda^2/(2M^2)})$,

$$\sup_{0 \leq k \leq n} || \sum_{k=0}^{n} R(k)|| \leq \lambda,$$

and hence $||T_4|| \leq b\lambda$.

Combining all upper bounds of the $T_i$, $i = 1, 2, 3, 4$, this gives with high probability

$$||\epsilon(t)|| \leq \Lambda \int_0^t ||\epsilon(t)||dt + t[\Lambda b(P + bN + M) + bN] + [2b^2N + b\lambda] \tag{12}$$

We could apply Gronwall's Lemma (Lemma 3), but we would get a lower quality bound[4].

[3] Recall that a sequence of random variables $Z_0, Z_1, \cdots$ is said to be martingale with respect to sequence $X_0, X_1, \cdots$ if, for all $n \geq 0$, we have (i) $Z_n$ is a function from $X_0, X_1, \cdots, X_n$ (ii) $\mathbb{E}[\,|Z_n|\,] < \infty$ (iii) $\mathbb{E}[\,Z_{n+1}\,|X_0, \cdots, X_n\,] = Z_n$. A function is martingale if it is a martingale with respect to itself.

[4] An additional $T$ factor in front of the exponential.

Let's do better: This gives for $v(t) = \int_0^t ||\epsilon(t)||$,

$$(v' - \Lambda v(t)) \exp(-\Lambda t) \leq t[\Lambda b(P + bN + M) + bN] \exp(-\Lambda t) + [2b^2 N + b\lambda] \exp(-\Lambda t)$$

Integrating, since $v(0) = 0$,

$$v(t) \exp(-\Lambda t) \leq \int_0^t [\Lambda b(P + bN + M) + bN] u \exp(-\Lambda u) du + \int_0^t [2b^2 N + b\lambda] \exp(-\Lambda u) du$$

This yields,

$$v(t) \exp(-\Lambda t) \leq [\Lambda b(P + bN + M) + bN] \frac{1 - (1 + \Lambda t)e^{-\Lambda t}}{\Lambda^2} + [2b^2 N + b\lambda] \frac{1 - \exp(-\Lambda t)}{\Lambda}$$

Reporting in (12), we get

$$||\epsilon(t)|| \leq [b(P + bN + M) + bN](\exp(\Lambda t) - (1 + \Lambda t)) + [2b^2 N + b\lambda](\exp(\Lambda t) - 1)$$
$$+t[\Lambda b(P + bN + M) + bN] + [2b^2 N + b\lambda].$$

That is to say

$$||\epsilon(t)|| \leq b[P + bN + M + \frac{N}{\Lambda}](\exp(\Lambda t) - 1) + [2b^2 N + b\lambda] \exp(\Lambda t)].$$

that holds with probability greater than $1 - \mathcal{O}(\frac{t}{b} e^{-\lambda^2/(2M^2)})$.

Taking $t = kb$ and $\lambda$ so that $\mathcal{O}(ke^{-\lambda^2}) < 1 - \mu$ gives the expected result.

## 7 Applications: Time Bounds

We get the following consequence, that can be applied to all the dynamics discussed in Section 5.

**Theorem 3.** *Assume the ordinary differential equation* (2) *is converging towards some Nash equilibria $Q^*$ for some initial condition $Q(0)$. Let $T(\epsilon)$ be the time needed to converge to some point at distance $\epsilon$ from $Q^*$.*

*Then for all $\epsilon > 0$, the stochastic algorithm started with initial state $Q(0)$ will converge with high probability towards a state at distance $2\epsilon$ of Nash equilibrium $Q^*$.*

*This will hold in a number of rounds of order*

$$\frac{1}{\epsilon} T(\epsilon) \exp(\Lambda T(\epsilon)),$$

*if one takes $b$ of order*

$$\epsilon \exp(-\Lambda T(\epsilon)).$$

*Proof.* Choose $T = kb = T(\epsilon)$, so that $X(T)$ is guaranteed to be at distance $\epsilon$ from $Q^*$. Applying Theorem 2, taking $b$ so that $be^{\Lambda T}(A + bB) \leq \epsilon$ guarantees that $||Q(k) - Q^*|| \leq ||Q(k) - X(kb)|| + ||X(kb) - Q^*|| \leq 2\epsilon$. This corresponds to the statement.

Hence, the time required by the stochastic algorithm is formally lower bounded by the time of the corresponding ordinary differential equation.

Notice that conversely, Theorem 2 gives a kind of lower bound: convergence of the stochastic algorithm implies convergence of the associated ordinary differential equation, in closely related time.

**Corollary 1.** *Assume that for some initial condition $Q(0)$ the stochastic algorithm is converging towards some Nash equilibria $Q^*$ for some initial condition $Q(0)$. Let $N(\epsilon, b)$ be the number of rounds needed to converge to some point at distance $\epsilon$ from $Q^*$.*

*Then the associated solution of initial value problem (2) is at time $t = N(\epsilon, b)b$ in some point at distance less than $\epsilon + be^{\Lambda N(\epsilon,b)b}(A + bB)$ from $Q^*$.*

*Proof.* This is just bounding $||X(kb) - Q^*||$ by $||X(kb) - Q(k)|| + ||Q(k) - Q^*||$.

In other words,

**Corollary 2.** *Bounding the time of convergence of any stochastic algorithm in the considered class is exactly the problem of bounding the time of the associated ordinary differential equation (2).*

## 8 Time of Convergence of Symmetric 2 × 2 Games

In this section, we consider dynamics from Section 5.1, hence yielding to continuous ordinary differential equation of the form (4), over generic symmetric two-player game $2 \times 2$ in which each player has only 2 pure strategies.

Let $A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$ be the payoff matrix[5]. Since replicator dynamics are invariant under a local shift of payoffs, matrix $A$ can be transformed into $\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$ with $a_1 = a_{1,1} - a_{2,1}$ and $a_2 = a_{2,2} - a_{1,2}$ without altering dynamics and time of convergence.

Dynamics (4) can be rewritten here

$$\begin{aligned} \frac{dq_{1,1}}{dt} &= (a_1 q_{2,1} - a_2 q_{2,2}) q_{1,1} q_{1,2} & \qquad \frac{dq_{2,1}}{dt} &= (a_1 q_{1,1} - a_2 q_{1,2}) q_{2,1} q_{2,2} \\ \frac{dq_{1,2}}{dt} &= -\frac{dq_{1,1}}{dt} & \frac{dq_{2,2}}{dt} &= -\frac{dq_{2,1}}{dt} \end{aligned}$$

Following [26], we can classify games into four categories.

- **Category I where $a_1 a_2 < 0$, $a_1 < 0$ and $a_2 > 0$:** there is an unique pure Nash equilibrium where $q_{1,1} = q_{2,1} = 0$.
  We always have $(a_1 q_{2,1} - a_2 q_{2,2}) < -\delta$ where $\delta = min(-a_1, a_2)$. Indeed, from $q_{2,1} + q_{2,2} = 1$, we have $(a_1 q_{2,1} - a_2 q_{2,2}) = (a_1 + a_2) q_{2,1} - a_2$. If $a_2 \le -a_1$, then, $(a_1 + a_2) \le 0$ and $(a_1 q_{2,1} - a_2 q_{2,2}) \le -a_2$. If $a_2 > -a_1$, then, $(a_1 + a_2) > 0$ and $(a_1 q_{2,1} - a_2 q_{2,2}) < a_1$.
  Since $\frac{dq_{1,1}}{dt} = (a_1 q_{2,1} - a_2 q_{2,2}) q_{1,1}(1 - q_{1,1})$, we have $\frac{dq_{1,1}}{dt} < \delta(1 - q_{1,1}) q_{1,1}$ And, hence $q_{1,1}(t) < e^{-\delta t} q_{1,1}(0)$ as soon as $q_{1,1}(0) \neq 0$. Symmetrically, we have $q_{2,1}(t) < e^{-\delta t} q_{2,1}(0)$ as soon as $q_{2,1}(0) \neq 0$. Hence, time of convergence $T(\epsilon)$ of order of $\frac{1}{\delta} \ln(\frac{1}{\epsilon})$.

[5] In order to follow [26], we consider payoffs, that is to say the opposite of costs. Maximizing payoffs correspond to minimizing costs.

- **Category IV where $a_1a_2 < 0$, $a_1 > 0$ and $a_2 < 0$:** there is an unique pure Nash equilibrium where $q_{1,1} = q_{2,1} = 1$. Similar computations show that we always have $(a_1q_{2,1} - a_2q_{2,2}) > \delta$ where $\delta = min(a_1, -a_2)$, and that $\frac{dq_{1,2}}{dt} < \delta(1 - q_{1,2})q_{1,2}$, $\frac{dq_{2,2}}{dt} < \delta(1 - q_{2,2})q_{2,2}$, yieldings a time of convergence $T(\epsilon)$ towards pure Nash equilibrium of order of $\frac{1}{\delta}\ln(\frac{1}{\epsilon})$.
- **Category II where $a_1a_2 > 0$ and $a_1$, $a_2$ are both positive**: there is one mixed Nash Equilibrium where $q_{1,1} = \frac{a_2}{a_1+a_2}$ and $q_{2,1} = \frac{a_2}{a_1+a_2}$ and two pure Nash Equilibria. One has the profile $q_{1,1} = 0$ and $q_{2,1} = 0$. Second has the profile $q_{1,1} = 1$ and $q_{2,1} = 1$.
  Let $\lambda = \frac{a_2}{a_1+a_2} - q_{2,1}$. Since $(a_1q_{2,1} - a_2q_{2,2}) = (a_1 + a_2)q_{2,1} - a_2$, we have $(a_1q_{2,1} - a_2q_{2,2}) = -\lambda(a_1 + a_2)$, and $\frac{dq_{1,1}}{dt} = -\lambda(a_1 + a_2)q_{1,1}(1 - q_{1,1})$. If $\lambda > 0$, then $\frac{dq_{1,1}}{dt} < -(a_1 + a_2)\lambda q_{1,1} < 0$. If $\lambda < 0$, then $\frac{dq_{1,1}}{dt} > -(a_1 + a_2)\lambda q_{1,1} > 0$, yielding $\frac{d(1-q_{1,1})}{dt} < (a_1 + a_2)\lambda(1 - q_{1,1}) < 0$.
  Symmetrically, let $\gamma = \frac{a_1}{a_1+a_2} - q_{1,1}$. If $\gamma > 0$, then $\frac{dq_{2,1}}{dt} < -(a_1 + a_2)\gamma q_{2,1} < 0$. If $\gamma < 0$, then $\frac{d(1-q_{2,1})}{dt} < (a_1 + a_2)\gamma(1 - q_{2,1}) < 0$.
  Hence, when $\lambda > 0$, $q_{1,1}$ decreases, while for $\lambda < 0$, $q_{1,1}$ increases. Symmetrically for $q_{2,1}$.
  The mixed Nash equilibrium is hence a saddle point. Outside the exception of a single curve through this point, any starting point will lead to a dynamic going to the pure Equilibria. This happens at a speed faster $\exp(-\delta t)$, for $\delta = \min(|\lambda|, |\gamma|)$ computed at the starting point in the diagonal quarters (corresponding to mixed Nash equilibria), hence yielding a time of convergence $T(\epsilon)$ is of order of $\frac{1}{\delta}\ln(\frac{1}{\epsilon})$. Other quarters are left in finite time to reach one of this quarters: this happens in a time of order $\frac{1}{\delta}\ln(\frac{1}{\epsilon'})$ where $\delta = \min(|\lambda|, |\gamma|)$ is computed at the point where this diagonal quarters are reached, and $\epsilon'$ is the distance of this point to the mixed Nash equilibrium.
- **Category III where $a_1a_2 > 0$ and where $a_1$, $a_2$ are both negative**: there are two pure Nash Equilibria and one mixed Nash Equilibrium. Nash equilibria have profile $q_{1,1} = \frac{a_2}{a_1+a_2}$, $q_{2,1} = \frac{a_2}{a_1+a_2}$ for the first, $q_{1,1} = 1$, $q_{2,1} = 0$, for the second, and $q_{1,1} = 0$, $q_{2,1} = 1$ for the third. Dynamics can be studied as in Category II. The difference is now that diagonal quarters are left in finite time, and that anti-diagonal quarters lead to exponentially converging dynamics towards Nash equilibria.

We see from the above discussion, that, possibly outside some single curves, or some unstable points, the time of convergence of the ordinary differential equation (4) is of order $\mathcal{O}(\log\frac{1}{\epsilon})$.

It follows from previous theorems and corollaries, that the corresponding stochastic algorithm can be guaranteed to converge to a point at distance less than $\epsilon$ from a Nash equilibrium in a time polynomial in $\frac{1}{\epsilon}$, taking $b$ of order polynomial in $\frac{1}{\epsilon}$.

Notice that for general games, one may expect the rate of convergence to be related to the eigenvalues of (the linearization of) the dynamic close to Equilibria, according to the classical general theory of dynamical systems [12], and hence the convergence to be also polynomial in $\frac{1}{\epsilon}$ in neighborhoods of Nash equilibria. There is no reason this holds outside their bassins of attraction, and this assumes non-degenerated cases (for e.g. that eigen values are all distinct): see [12].

## References


1. M. Benaïm. Dynamics of stochastic approximation algorithms. *Séminaire de Probabilités XXXIII. Lecture Notes in Math*, 1709:1–68, 1999.
2. Petra Berenbrink and Oliver Schulte. Evolutionary equilibrium in bayesian routing games: Specialization and niche formation. In *ESA*, pages 29–40, 2007.

3. Steve Chien and Alistair Sinclair. Convergence to approximate nash equilibria in congestion games. In *SODA*, pages 169–178, 2007.
4. George Christodoulou, Elias Koutsoupias, and Akash Nanavati. Coordination mechanisms. In *Automata, Languages and Programming: 31st International Colloquium*, volume 3142 of *Lecture Notes in Computer Science*, pages 345–357. Springer, 2004.
5. F. Comets and T. Meyre. *Calcul stochastique et modeles de diffusions*. Dunod Paris, 2006.
6. Eyal Even-Dar, Alexander Kesselman, and Yishay Mansour. Convergence time to Nash equilibrium in load balancing. *ACM Transactions on Algorithms*, 3(3), 2007.
7. S. Fischer, H. Räcke, and B. Vöcking. Fast convergence to Wardrop equilibria by adaptive sampling methods. *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 653–662, 2006.
8. S. Fischer and B. Vocking. On the Evolution of Selfish Routing. *Algorithms–ESA 2004: 12th Annual European Symposium, Bergen, Norway, September 14-17, 2004, Proceedings*, 2004.
9. Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*. Number 624. December 1996. available at http://ideas.repec.org/p/cla/levarc/624.html.
10. Paul W. Goldberg. Bounds for the convergence rate of randomized local search in a multiplayer load-balancing game. In *PODC '04: Proceedings of the twenty-third annual ACM symposium on Principles of distributed computing*, pages 131–140, New York, NY, USA, 2004. ACM.
11. C. Harris. On the Rate of Convergence of Continuous-Time Fictitious Play. *Games and Economic Behavior*, 22(2):238–259, 1998.
12. Morris W. Hirsch, Stephen Smale, and Robert Devaney. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Elsevier Academic Press, 2003.
13. J. Hofbauer and K. Sigmund. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 4:479–519, 2003.
14. D.S. Johnson, C.H. Papadimitriou, and M. Yannakakis. How easy is local search? *Journal of Computer System Science*, 37(1):79–100, 1988.
15. E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Symposium on Theoretical Computer Science (STACS'99)*, pages 404–413, Trier, Germany, 4–6March 1999.
16. M.A.L. Thathachar K.S. Narendra. Learning automata: An introduction. 1989.
17. H.J. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*. Springer, 2003.
18. L. Libman and A. Orda. Atomic Resource Sharing in Noncooperative Networks. *Telecommunication Systems*, 17(4):385–409, 2001.
19. Michael Mitzenmacher and Eli Upfal. *Probability and Computing. Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.
20. D. Monderer and L.S. Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.
21. A. Orda, R. Rom, and N. Shimkin. Competitive routing in multiuser communication networks. *IEEE/ACM Transactions on Networking (TON)*, 1(5):510–521, 1993.
22. M.A.L. Thathachar P.S. Sastry, V.V. Phansalkar. Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information. *IEEE transactions on system, man, and cybernetics*, 24(5), 1994.
23. R.W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.
24. W.H. Sandholm. Potential games with continuous player sets. *Journal of Economic Theory*, 97(1):81–108, 2001.
25. D.W. Stroock and SRS Varadhan. *Multidimensional Diffusion Processes*. Springer, 1979.
26. Jörgen W. Weibull. *Evolutionary Game Theory*. The MIT Press, 1995.

## A A General Theorem about Approximation of Diffusions

We will use the following theorem from [25, theorem 11.2.3]. The following presentation is inspired by the presentation of it in [5, Theorem 5.8, page 96].

Suppose that for all integers $b > 0$, we have an homogeneous Markov chain $(Y_k^{(b)})$ in $\mathbb{R}^d$ with transition kernel $\pi^{(b)}(x, dy)$, meaning that the law of $Y_{k+1}^{(b)}$, conditioned on $Y_0^{(b)}, \cdots, Y_k^{(b)}$, depends only on $Y_k^{(b)}$ and is given, for all Borel set $B$, by $P(Y_{k+1}^{(b)} \in B | Y_k^{(b)}) = \pi^{(b)}(Y_k^{(b)}, B)$, almost surely.

Define for $x \in \mathbb{R}^d$,

$$\begin{aligned}
d^{(b)}(x) &= \frac{1}{b} \int (y - x) \pi^{(b)}(x, dy), \\
a^{(b)}(x) &= \frac{1}{b} \int (y - x)(y - x)^* \pi^{(b)}(x, dy), \\
K^{(b)}(x) &= \frac{1}{b} \int (y - x)^3 \pi^{(b)}(x, dy), \\
\Delta_\epsilon^{(b)}(x) &= \frac{1}{b} \pi^{(b)}(x, B(x, \epsilon)^c),
\end{aligned}$$

where $B(x, \epsilon)^c$ denotes the complement of the ball with radius $\epsilon$, centered at $x$.

The coefficients $d^{(b)}$ and $a^{(b)}$ can be interpreted as the instantaneous drift and the variance (or matrix of covariance) of $X^{(b)}$.

Define

$$X^{(b)}(t) = Y_{\lfloor t/b \rfloor}^{(b)} + (t/b - \lfloor t/b \rfloor)(Y_{\lfloor t/b+1 \rfloor}^{(b)} - Y_{\lfloor t/b \rfloor}^{(b)}).$$

**Theorem 4 ([25, theorem 11.2.3], [5, Theorem 5.8, page 96]).** *Suppose that there exist some continuous functions $d, b$, such that for all $R < +\infty$,*

$$\begin{aligned}
\lim_{b \to 0} sup_{|x| \le R} |a^{(b)}(x) - a(x)| &= 0 \\
\lim_{b \to 0} sup_{|x| \le R} |d^{(b)}(x) - d(x)| &= 0 \\
\lim_{b \to 0} sup_{|x| \le R} \Delta_\epsilon^{(b)} &= 0, \forall \epsilon > 0 \\
\sup_{|x| \le R} K^{(b)}(x) &< \infty.
\end{aligned}$$

*With $\sigma$ a matrix such that $\sigma(x)\sigma^*(x) = a(x)$, $x \in \mathbb{R}^d$, we suppose that the stochastic differential equation*

$$dX(t) = d(X(t))dt + \sigma(X(t))dB(t), \qquad X(0) = x, \tag{13}$$

*has a unique weak solution for all $x$. This is in particular the case, if it admits a unique strong solution.*

*Then for all sequences of initial conditions $Y_0^{(b)} \to x$, the sequence of random processes $X^{(b)}$* weakly converges *to the diffusion given by Equation* (13). *In other words, for all functions $F : \mathcal{C}(\mathbb{R}^+, \mathbb{R}) \to \mathbb{R}$ bounded and continuous, one has*

$$\lim_{b \to 0} E[F(X^{(b)})] = E[F(X)].$$

## B Proof of Theorem 1

Theorem 1 follows from previous theorem. Consider $(Y_k^{(b)})$ to be

$$Y_k^{(b)} = (Q(k))$$

with the corresponding $b$, which is indeed an homogeneous Markov chain. Let $\pi^{(b)}(Q, dy)$ be its transition kernel.

We have

$$\begin{aligned} d_i^{(b)}(Q) &= \tfrac{1}{b}\int (y_i - q_i)\pi^{(b)}(x, dy) \\ &= \tfrac{1}{b}\mathbb{E}[\ \Delta q_i\ |Q\ ] \\ &= \tfrac{b}{b}\tilde{F}_i^b(Q) \\ &\rightarrow G_i(Q) \qquad\quad \text{when } b \rightarrow 0 \end{aligned}$$

and

$$\begin{aligned} a_{i,j}^{(b)}(Q) &= \tfrac{1}{b}\int (y_i - q_i)(y_j - q_j)^*\pi^{(n)}(x, dy) \\ &= \tfrac{b^2}{b}\mathbb{E}[\ \Delta q_i \Delta q_j\ |Q\ ] \\ &= \mathcal{O}(b) \\ &\rightarrow 0 \text{ when } b \rightarrow 0 \end{aligned}$$

In the same vein, clearly $K^{(b)}(Q)$ stay bounded, being in $\mathcal{O}(b^2)$.

Now, from the fact that compact $K$ must be kept invariant by the dynamics, $F_i^b(.)$ must have a compact support. This means that $\pi^{(b)}(Q, B(Q,\epsilon)^c)$ is 0 for $b$ sufficiently small. Hence $\lim_{b\rightarrow 0} sup_{|x|\leq R}\Delta_\epsilon^{(b)} = 0$, $\forall \epsilon > 0$.

Hence, we have all the hypotheses of previous theorem with $a(Q) = 0$ and

$$d(Q) = G(Q)$$

observing that the corresponding stochastic differential equation $dQ(t) = d(Q(t))dt + \sigma(Q(t))dB(t)$ turns out to be an ordinary differential equation, whose solution is unique by (classical) Cauchy Lipschtiz theorem.

## C Proof of Proposition 1

The following are well-known (and obtained by just playing with definitions).

**Lemma 1.** *A strategy profile $Q$ is a Nash Equilibrium iff $c_i(q_i, Q_{-i}) \leq c_i(e_\ell, Q_{-i})$ for all $1 \leq i \leq n$, $1 \leq \ell \leq m_i$.*

**Corollary 3.** *In a Nash Equilibrium, we have $c_i(q_i, Q_{-i}) = c_i(e_\ell, Q_{-i})$ for all $1 \leq i \leq n$, $1 \leq \ell \leq m_i$ with $q_{i,\ell} > 0$.*

Proposition 1 is then an instance of the so-called folk-theorems of Evolutionary Game Theory [13]. For completeness, the proof goes as follows: From Corollary 3, clearly any Nash equilibria must also vanish the right-hand side of Equation (4).

A non-Nash equilibrium $Q$ is not stable: Indeed, if $Q$ is not a Nash equilibrium, this means that for some $i$, and some $\ell$ we have $c_i(q_i, Q_{-i}) > c_i(e_\ell, Q_{-i})$. By bilinearity and continuity of $c_i$, function $c_i(q_i - e_\ell, Q_{-i})$ must be strictly positive (say greater than $\epsilon$) on some neighborhood of $Q$.

On this neighborhood, $\frac{dq_{i,\ell}}{dt}$ is greater than $p_i q_{i,\ell}\epsilon$, and hence the point is left exponentially faster (faster than exponential $q_{i,\ell}(0)\exp(p_i \epsilon t)$).

In a corner of $K$, we have for all $i$, $q_i = e_\ell$ for some $\ell$. Then clearly $q_{i,\ell'} = 0$ for index $\ell' \neq \ell$, and $c_i(e_\ell, Q_{-i}) - c_i(q_i, Q_{-i}) = 0$ for index $\ell' = \ell'$. Hence, the right-hand side of Equation (4) is always null, and hence any corner is a stationary point.

More generally any state $Q$ in which all strategies in its support perform equally well, is clearly a stationary point from the definition of the dynamic.

## D Azuma-Hoeffding's Inequality

**Lemma 2 (Azuma-Hoeffding's Inequality: see e.g. [19]).** *Let $Z_1, Z_2, \cdots, Z_n$ a martingale such that*

$$|Z_k - Z_{k-1}| \leq c_k.$$

*Then for all $t \geq 0$ and all $\lambda > 0$,*

$$Pr(|Z_t - Z_0| \geq \lambda) \leq 2e^{-\lambda^2/(2\sum_{k=1}^{t} c_k^2)}.$$

## E Gronwall's Lemma

**Lemma 3 (Gronwall's Lemma).** *. Let $u(t)$ and $g(t)$ be non-negative continuous functions on $I = [0, \infty)$ for which the inequality*

$$u(t) \leq c + \int_a^t g(s)u(s)ds,$$

*for $t \in I$. Then*

$$u(t) \leq c\exp(\int_a^t g(s)ds),$$

*for $t \in I$.*