extended abstract

Joint Strategy Fictitious Play with Inertia for Potential Games

Jason R. Marden
University of California Los Angeles
marden@seas.ucla.edu

Gürdal Arslan
University of Hawaii
gurdal@hawaii.edu

Jeff S. Shamma
University of California Los Angeles
shamma@ucla.edu

1 Overview

We consider finite multi-player repeated games involving a large number of players with large strategy spaces and enmeshed utility structures. In these "large-scale" games, players are inherently faced with limitations in both their observational and computational capabilities. Accordingly, players in large-scale games need to make their decisions using algorithms that accommodate limitations in information gathering and processing. A motivating example is a congestion game [Ros73] in a complex transportation system [BAL85], in which a large number of vehicles make daily routing decisions to optimize their own objectives in response to their observations. In this setting, observing and responding to the individual actions of all vehicles on a daily basis would be a formidable task for any individual driver. This disqualifies some of the well known decision making models such as "Fictitious Play" (FP) [FL98] as suitable models for driver routing behavior. A more realistic assumption on the information tracked and processed by an individual driver is the daily aggregate congestion on the specific roads that are of interest to that driver. We will show that Joint Strategy Fictitious Play (JSFP) [FL98, FK93, MS97], a close variant of FP, when modified to include some inertia, accommodates such information aggregation. We establish the convergence of JSFP with inertia to a pure Nash equilibrium in finite potential games, in both cases of averaged or exponentially discounted historical data.

It turns out that there is a strong similarity between the JSFP discussed herein and the regret matching algorithm [HMC00]. A player's regret for a particular choice is defined as the difference between 1) the utility that would have been received if that particular choice was played for all the previous stages and 2) the average utility actually received in the previous stages. A player using the regret matching algorithm updates a regret vector for each possible choice, and selects actions according to a probability proportional to positive regret. Similarly, a player using the JSFP algorithm evaluates an average reward for each possible action. The average reward of a particular choice is the utility that would have been received if that particular choice was played for all previous stages. A player would than select an action with the highest average reward. It turns out that this is equivalent to playing an action that yielded the highest regret.

A current open question is whether player choices would converge in coordination-type games when all players use the regret matching algorithm (except for the special case of two-player games [HMC03]). There are finite memory versions of the regret matching algorithm and various generalizations [You05], such as playing best or better responses to regret over the last m stages, that are proven to be convergent in weakly acyclic games when players use some sort of inertia. These finite memory algorithms do not require each player to track the behavior of other players individually. Rather, each player needs to remember the utilities actually received and the utilities that could have been received in the last m stages. In contrast, a player using JSFP best responds according to accumulated experience over the entire history by using a simple recursion which can also incorporate exponential discounting of the historical data.

This talk presents an analysis of the convergence properties of JSFP with inertia for generalized ordinal potential games. A congestion game simulation will be presented to illustrate the computational effectiveness of this algorithm.

2 Setup

Consider a finite game with n-player set $\mathcal{P} := \{\mathcal{P}_1, ..., \mathcal{P}_n\}$ where each player $\mathcal{P}_i \in \mathcal{P}$ has an action set Y_i and a utility function $U_i : Y \to \mathbb{R}$ where $Y = Y_1 \times ... \times Y_n$. We will frequently write $U(y)$ as $U(y_i, y_{-i})$ where y_{-i} denotes the profile of the actions of the players *other than* player $\mathcal{P}_i$.

Definition 2.1 (Nash Equilibrium) *An action profile y^* is called a pure Nash equilibrium if for all players $\mathcal{P}_i \in \mathcal{P}$,*

$$U_i(y_i^*, y_{-i}^*) = \max_{y_i \in Y_i} U_i(y_i, y_{-i}^*). \tag{1}$$

We will consider "generalized ordinal potential games", defined as follows.

Definition 2.2 (Generalized Ordinal Potential Games) *A finite n-player game with action sets $\{Y_i\}_{i=1}^n$ and utility functions $\{U_i\}_{i=1}^n$ is a **generalized ordinal potential game** if, for some potential function $\phi : Y_1 \times ... \times Y_n \to \mathbb{R}$,*

$$U_i(y_i', y_{-i}) - U_i(y_i'', y_{-i}) > 0 \Rightarrow \phi(y_i', y_{-i}) - \phi(y_i'', y_{-i}) > 0,$$

for every player, and for every $y_{-i} \in \times_{j \neq i} Y_j$ and for every $y_i', y_i'' \in Y_i$.

2.1 Fictitious Play

We start with the well-known Fictitious Play (FP) process [FL98].

Define the *empirical frequency*, $q_i^{\bar{y}_i}(t)$, as the percentage of stages at which player $\mathcal{P}_i$ has chosen the action $\bar{y}_i \in Y_i$ up to time $t-1$, i.e.,

$$q_i^{\bar{y}_i}(t) = \frac{1}{t}\sum_{\tau=0}^{t-1} I\{y_i(\tau) = \bar{y}_i\},$$

where $y_i(k) \in Y_i$ is player $\mathcal{P}_i$'s action at time k and $I\{\cdot\}$ is the indicator function. Let $q_i(t)$ denote the empirical frequency vector for player $\mathcal{P}_i$ formed by the components $\{q_i^{y_i}(t)\}_{y_i \in Y_i}$. Note that the dimension of $q_i(t)$ is the cardinality of $|Y_i|$.

The action of player $\mathcal{P}_i$ at time t is based on the (incorrect) presumption that other players are playing *randomly* and *independently* according to their empirical frequencies. Under this presumption, the expected utility for the action $\bar{y}_i \in Y_i$ is

$$U_i(\bar{y}_i, q_{-i}(t)) := \sum_{y_{-i} \in Y_{-i}} U_i(\bar{y}_i, y_{-i}) \prod_{j \neq i} q_j^{y_j}(t), \tag{2}$$

where $q_{-i}(t) := \{q_1(t), ..., q_{i-1}(t), q_{i+1}(t), ..., q_n(t)\}$ and $Y_{-i} := \times_{j \neq i} Y_j$. In the FP process, player $\mathcal{P}_i$ uses this expected utility by selecting an action at time t from the set

$$BR_i(q_{-i}(t)) := \{\tilde{y}_i \in Y_i : U_i(\tilde{y}_i, q_{-i}(t)) = \max_{y_i \in Y_i} U_i(y_i, q_{-i}(t))\}.$$

The set $BR_i(q_{-i}(t))$ is called player $\mathcal{P}_i$'s best response to $q_{-i}(t)$. In case of a non-unique best response, player $\mathcal{P}_i$ makes a random selection from $BR_i(q_{-i}(t))$.

It is known that the empirical frequencies generated by FP converge to a Nash equilibrium in potential games [MS96].

Note that FP as describe above requires each player to observe the actions made by every other individual player. Moreover, choosing an action based on the predictions (2) amounts to enumerating all possible joint actions in $\times_j Y_j$ every stage for each player. Hence, FP is computationally prohibitive as a decision making model in large-scale games.

2.2 Joint Strategy Fictitious Play

In JSFP, each player tracks the empirical frequencies of the *joint actions* of all other players. In contrast to FP, the action of player $\mathcal{P}_i$ at time t is based on the (still incorrect) presumption that other players are playing *randomly* but *jointly* according to their *joint* empirical frequencies, i.e., each player views all other players as a collective group.

Let $z_{-i}^{\bar{y}_{-i}}(t)$ be the percentage of stages at which players other then player $\mathcal{P}_i$ have chosen the joint action profile $\bar{y}_{-i} \in Y_{-i}$ up to time $t-1$, i.e.,

$$z_{-i}^{\bar{y}_{-i}}(t) = \frac{1}{t}\sum_{\tau=0}^{t-1} I\{y_{-i}(\tau) = \bar{y}_{-i}\}. \tag{3}$$

Let $z_{-i}(t)$ denote the empirical frequency vector formed by the components $\{z_{-i}^{\bar{y}_{-i}}(t)\}_{\bar{y}_{-i} \in Y_{-i}}$. Note that the dimension of $z_{-i}(t)$ is the cardinality $|\times_{i \neq j} Y_j|$.

Similarly to FP, player $\mathcal{P}_i$'s action at time t is based on an expected utility for the action $\bar{y}_i \in Y_i$, but now based on the joint action model of opponents given by[1]

$$U_i(\bar{y}_i, z_{-i}(t)) := \sum_{y_{-i} \in Y_{-i}} U_i(\bar{y}_i, y_{-i}) z_{-i}^{y_{-i}}(t). \tag{4}$$

In the JSFP process, player $\mathcal{P}_i$ uses this expected utility by selecting an action at time t from the set

$$BR_i(z_{-i}(t)) := \{\tilde{y}_i \in Y_i : U_i(\tilde{y}_i, z_{-i}(t)) = \max_{y_i \in Y_i} U_i(y_i, z_{-i}(t))\}.$$

When written in this form, JSFP appears to have a computational burden for each player that is even higher than that of FP, since tracking the empirical frequencies $z_{-i}(t) \in \Delta(Y_{-i})$ of the joint actions of the other players is more demanding for player $\mathcal{P}_i$ than tracking the empirical frequencies $q_{-i}(t) \in \times_{j \neq i} \Delta(Y_j)$ of the actions of the other players individually, where $\Delta(Y)$ denotes the set of probability distributions on a finite set Y. However, it is possible to rewrite JSFP to significantly reduce the computational burden on each player.

To choose an action at any time t, player $\mathcal{P}_i$ using JSFP needs only the predicted utilities $U_i(\bar{y}_i, z_{-i}(t))$ for each $\bar{y}_i \in Y_i$. Substituting (3) into (4) results in

$$V_i^{\bar{y}_i}(t) := U_i(\bar{y}_i, z_{-i}(t)) = \frac{1}{t}\sum_{\tau=0}^{t-1} U_i(\bar{y}_i, y_{-i}(\tau)),$$

[1] Note that we use the same notation for the related quantities $U(y_i, y_{-i})$, $U(y_i, q_{-i})$, and $U(y_i, z_{-i})$, where the latter two are derived from the first as defined in equations (2) and (4), respectively.

which is the average utility player $\mathcal{P}_i$ would have received if action $\bar{y}_i$ had been chosen at every stage up to time $t-1$ and other players used the same actions. Furthermore, this is the same (hypothetical) average utility that is used in the aforementioned no-regret algorithms.

The average utility $V_i^{\bar{y}_i}(t)$ admits the following simple recursion,

$$V_i^{\bar{y}_i}(t+1) = \frac{t}{t+1} V_i^{\bar{y}_i}(t) + \frac{1}{t+1} U_i(\bar{y}_i, y_{-i}(t)),$$

which permits the JSFP dynamics to proceed without requiring each player to track the empirical frequencies of the joint actions of the other players and without requiring each player to compute an expectation over the space of the joint actions of all other players. Each player using JSFP merely updates the predicted utilities for each action using the recursion above, and chooses an action each stage with maximal predicted utility.

The convergence properties, even for potential games, of JSFP in the case of more than two players is unresolved.[2] We will establish convergence of JSFP in the case where players use some inertia, i.e., players are hesitant to change actions even when there is a perceived opportunity for improvement.

2.3 Joint Strategy Fictitious Play with Inertia

The JSFP with inertia process is defined as follows. Players choose their actions according to the following rules:

JSFP-1: If the action $y_i(t-1)$ chosen by player $\mathcal{P}_i$ at time $t-1$ belongs to $BR_i(z_{-i}(t))$, then $y_i(t) = y_i(t-1)$.

JSFP-2: Otherwise, player $\mathcal{P}_i$ chooses an action, $y_i(t)$, at time t according to the probability distribution

$$\alpha_i(t)\beta_i(t) + (1-\alpha_i(t))\mathbf{v}^{y_i(t-1)},$$

where $\alpha_i(t)$ is a parameter representing player $\mathcal{P}_i$'s willingness to optimize at time t, $\beta_i(t) \in \Delta(Y_i)$ is any probability distribution whose support is contained in the set $BR_i(z_{-i}(t))$, and $\mathbf{v}^{y_i(t-1)}$ is the vertex of $\Delta(Y_i)$ corresponding to the action $y_i(t-1)$.

According to these rules, player $\mathcal{P}_i$ will stay with the previous action $y_i(t-1)$ with probability $1-\alpha_i(t)$ even when there is a perceived opportunity for utility improvement. We make the following standing assumption on the players' willingness to optimize.

Assumption 2.1 *There exist constants $\underline{\varepsilon}$ and $\bar{\varepsilon}$ such that for all time $t \geq 0$ and for all players $i \in \{1, ..., n\}$,*

$$0 < \underline{\varepsilon} < \alpha_i(t) < \bar{\varepsilon} < 1.$$

This assumption implies that players are always willing to optimize with some nonzero inertia[3]

2.3.1 Convergence to Nash Equilibrium

We will assume that no player is indifferent between distinct action profiles.

Assumption 2.2 *Player utilities satisfy*

$$U_i(y_i^1, y_{-i}) \neq U_i(y_i^2, y_{-i}),\ \forall\, y_i^1, y_i^2 \in Y_i,\ y_i^1 \neq y_i^2,\ \forall\, y_{-i} \in Y_{-i},\ \forall\, i \in \{1, ..., n\}. \tag{5}$$

The following establishes the main result regarding the convergence of JSFP with inertia.

Theorem 2.1 *In any finite generalized ordinal potential game in which no player is indifferent between distinct strategies as in Assumption 2.2, the action profiles $y(t)$ generated by JSFP with inertia under Assumption 2.1 converge to a pure Nash equilibrium almost surely.*

2.4 Illustrative Simulations

The talk will present simulations of the JSFP with Inertia algorithm applied to congestion games.

References

[BAL85] M. Ben-Akiva and S. Lerman. *Discrete-Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA, 1985.

[FK93] D. Fudenberg and D. Kreps. Learning mixed equilibria. *Games and Economic Behavior*, **5**(3):320–367, 1993.

[FL98] D. Fudenberg and D. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1998.

[HMC00] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, **68**(5):1127–1150, 2000.

[HMC03] S. Hart and A. Mas-Colell. Regret based continuous-time dynamics. *Games and Economic Behavior*, **45**:375–394, 2003.

[MS96] D. Monderer and L. Shapley. Potential games. *Games and Economic Behavior*, **14**:124–143, 1996.

[MS97] D. Monderer and A. Sela. Fictitious play and no-cycling conditions. Technical report, 1997.

[Ros73] R. W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *Int. J. Game Theory*, **2**:65–67, 1973.

[You05] H. P. Young. *Strategic Learning And Its Limits*. Oxford University Press, 2005.

[2] For two player games, JSFP and standard FP are equivalent, hence the convergence results for FP hold for JSFP.

[3] This assumption can be relaxed to holding for sufficiently large t, as opposed to all t.