Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Joint Strategy Fictitious Play with Inertia for Potential Games

Jason R. Marden, Gürdal Arslan and Jeff S. Shamma

***Abstract*— 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 in a complex transportation system, 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) 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), a close variant of FP, accommodates such information aggregation. Furthermore, we establish the convergence of JSFP to a pure Nash equilibrium in congestion games, or equivalently in finite potential games, when players use some inertia in their decisions and in both cases of with or without exponential discounting of the historical data.**

I. OVERVIEW

We consider finite multi-player repeated games involving a large number of players with large strategy spaces and enmeshed utility structures. In these so called "large-scale" games, players are inherently faced with limitations in both their observational and computational capabilities. Accordingly, players in such large-scale games need to make their decisions using algorithms that accommodate limitations in information gathering and processing. The main objective of this paper is to study the convergence properties of a particular algorithm, called Joint Strategy Fictitious Play (JSFP) [1], [2], [3], which, we will argue, is a plausible decision making model for certain large-scale games.

Our motivating example of a large-scale game is a congestion game [4] in a complex transportation system [5] in which a large number of vehicles make daily routing decisions to optimize their own objectives in response to their own 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) [1] as appropriate models for driver travel choice behavior. A more realistic measure on the information tracked and processed by an individual driver is the daily aggregate congestion on the roads that are of interest to driver [6]. It turns out that JSFP, a close variant of FP, accommodates such information aggregation.

Deriving accurate driver behavior models has been an active research area in transportation science [6], [7], [8]. Consider the general framework for the day-to-day adjustment process proposed in [6], where drivers base their daily routing decisions on their perception of the traffic conditions. Drivers update their perceptions on a daily basis using historical information on the traffic network conditions. This includes a driver's own daily personal experience on the chosen routes, as well as aggregate information on daily traffic conditions for alternative routes. Drivers may acquire information on traffic conditions via the internet, media traffic reports, or even communications with friends. The framework proposed in [6] also incorporates pre-trip predicted travel information provided by a Driver Information System. Finally, drivers are assumed to select their travel patterns using a random utility model, where driver utilities depend on their perceptions.

In this paper, we consider a specification of the above dynamics in which each drivers keep track of the (weighted) average utility they "would have" received on their available routes. A driver's utility for an available route reflects the congestion experienced on that route. Every day, each driver selects the route that would have yielded the highest average utility in previous days. At the end of the day, drivers update their perceptions of the congestion on available routes using only aggregate information of the traffic conditions.

We will show that such driver behavior can be modeled as JSFP, even though JSFP, if taken at face value, seemingly has a significantly more demanding information gathering and processing requirement for each driver than suggested. The main result is to establish almost sure convergence of JSFP to a pure Nash equilibrium in congestion games, or equivalently in finite potential games [9], when drivers use some inertia in decisions, i.e., a reluctance to change the daily route. Such inertia is also motivated by the framework of [6].

Convergence of driver behavior to a steady-state regime is of interest in its own right. See [6], [7] where the issue

Research supported by NSF grant #CMS–0339228 and ARO grant #W911NF–04–1–0316

J. R. Marden is with the Department of Mechanical and Aerospace Engineering, University of California, 46-148 Engineering IV, Los Angeles, CA 90095-1597, marden@seas.ucla.edu. Mr. Marden would like to thank Raytheon Systems Corporation for their support.

G. Arslan is with the Department of Electrical Engineering, University of Hawaii at Manoa, 440 Holmes Hall, 2540 Dole Street, Honolulu, HI 96822, gurdal@hawaii.edu.

J. S. Shamma is with the Department of Mechanical and Aerospace Engineering, University of California, 37-146 Engineering IV, Los Angeles, CA 90095-1597, shamma@seas.ucla.edu.

of convergence of driver behavior is raised. The significance of our convergence result is more relevant in the case where driver utilities are engineered, perhaps indirectly, in such a way that the resulting game is a coordination type, and a pure Nash equilibrium is the desired operating point. For instance, in a traffic congestion game, a global planner may render a particular pure Nash equilibrium the optimal solution from the planner's own perspective by taxation or other means [10]. Then, the question would become whether the drivers that are using a reasonable decision making rule such as JSFP would settle at the optimal Nash equilibrium.

We will now review some of the well-known decision making models and discuss their limitations in large-scale games, or specifically in traffic congestion games. See the monographs [1], [11], [12], [13], [14] and survey article [15] for a more comprehensive review.

The well-known FP algorithm requires that each player views all other players as independent decision makers, [9]. In the FP framework, each player observes the decisions made by all other players and computes the empirical frequencies (i.e, running averages) of these observed decisions. Then, each player best responds to the empirical frequencies of other players' decisions by first computing the expected utility for each strategy choice under the assumption that the other players will independently make their decisions probabilistically according to the observed empirical frequencies. FP is known to be convergent to a Nash equilibrium in potential games, but need not converge for other classes of games. General convergence issues are discussed in [16], [17], [18].

A recent paper [19] introduced a version of FP, called sampled FP, in order to avoid computing an expected utility based on the empirical frequencies. In sampled FP, each player selects samples from the strategy space of every other player according to the empirical frequencies of that player's past decisions. Each player than computes an average utility for each strategy choice based off of these samples. Each player still has to observe the decisions made by all other players to compute the empirical frequencies of these observed decisions. Furthermore, sampled FP is proved to be convergent only in identical interest games, and the number of samples needed to guarantee convergence grows unboundedly.

Let us now review some of the learning algorithms that are convergent in a large class of coordination games called "weakly acyclic" games [11]. In the best-reply and better-reply dynamics of [20], each player plays a best or better response to the most recent decision profile whenever that player randomly receives a strategy revision opportunity. Players using such best-reply or better-reply dynamics cannot have any memory. In a traffic congestion problem this lack of memory translates to the players remembering only the most recent congestion information, which is an unsuitable model since drivers typically rely on past information for future decision making [6]. Another learning algorithm convergent in weakly acyclic games is adaptive play [21], [11] where players have finite recall. However, adaptive play requires each player to track the individual behavior of all other players for recall windows greater than one. Thus, as the size of player memory grows, adaptive play suffers from the same computational setback as FP.

It turns out that there is a strong similarity between JSFP and the regret matching algorithm [22]. 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 steps and 2) the average utility actually received in the previous steps. 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. One can view JSFP as an algorithm where each player also best responds to regret. In particular, a player chooses according to maximum regret. To the authors' knowledge, it is not known whether player choices would converge in coordination-type games when all players use the regret matching algorithm (except for two-player games [23]). However, finite memory versions of the regret matching algorithm and its generalizations [12], such as playing best or better responses to regret over the last m time steps, are proven to be convergent in weakly acyclic games when players use some inertia. These finite memory algorithms do not require each player to track the behavior of other players individually. However, each player still needs to remember the utilities actually received and the utilities that could have been received in the last m time steps. In contrast, a player using JSFP best responds to regret computed over the entire history by using a simple recursion which can also incorporate exponential discounting of the historical data.

There are also payoff based dynamics, where each player observes only the actual utilities received and uses a Reinforcement Learning (RL) algorithm [24], [25] to make future choices. In the context of the traffic congestion game, a driver using an RL-like algorithm would observe only the congestion personally experienced on routes actually travelled. Therefore, a driver using an RL-like algorithm cannot make use of the information on alternate routes that could be available through means other than personal experience. Convergence of player choices when all players use an RL-like algorithm is proved only in identical interest games [26], [27], [28], [29] but often with delicate tuning of various parameters, e.g., having each player adapt at a different time scale. Finally, the payoff based dynamics with finite-memory presented in [29] leads to a Pareto-optimal outcome in generic common interest games.

The remainder of the paper is organized as follows. Section 2 is devoted to decision making in congestion games. Section 3 presents JSFP. Convergence of JSFP process with inertia in potential games is presented in Section 4. Section 5 studies convergence when historical information is discounted exponentially. Lastly, Section 6 presents an illustrative example.

II. Decision Making in Congestion Games

We consider a transportation network with a finite set R of road segments that needs to be shared by a set of selfish drivers labelled as $D := \{d_1, ..., d_n\}$. Each driver has a fixed origin/destination pair connected through multiple routes. The set of all routes available to driver d_i is denoted by Y_i. A typical route y_i in Y_i consists of multiple road segments, therefore, $y_i \subset R$. Driver d_i taking route y_i incurs a cost c_r for each road segment $r \in y_i$. The utility of driver d_i taking route y_i is defined as the negative of the total cost incurred, i.e., $U_i = -\sum_{r \in y_i} c_r$.

The setup described above leads to a congestion game [4], because we assume that the cost incurred in a road segment depends on the total number of drivers sharing that road. Hence, the utility of each driver will depend on the routes chosen by other drivers. The utility of driver d_i is stated more precisely as

$$U_i(y) = -\sum_{r \in y_i} c_r(\sigma_r(y)),$$

where $y := (y_1, ..., y_n)$ is the profile of routes chosen by all drivers and $\sigma_r(y)$ is the total number of drivers using the road segment r.

A class of games closely related to congestion games is potential games. A finite n-player game with choice sets $\{Y_i\}_{i=1}^n$ and utility functions $\{U_i\}_{i=1}^n$ constitutes a potential game if, for some potential function $\phi : \times_j Y_j \mapsto \mathcal{R}$,

$$U_i(y_i', y_{-i}) - U_i(y_i'', y_{-i}) = \phi(y_i', y_{-i}) - \phi(y_i'', y_{-i}),$$

for every player, for every $y_{-i} \in \times_{j \neq i} Y_j$, and for every $a_i', a_i'' \in Y_i$. It turns out that every congestion game is a potential game and every finite potential game is isomorphic to a congestion game [9].

Let y_{-i} denote the profile of the route choices of the drivers *other than* driver d_i, i.e.,

$$y_{-i} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n).$$

With this notation, we will sometimes write a profile y of route choices as (y_i, y_{-i}). Similarly, we may write $U_i(y)$ as $U_i(y_i, y_{-i})$.

A profile y^* of route choices is called a *pure Nash equilibrium*[1] if, for all drivers $d_i \in D$,

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

Therefore, if the routes chosen by the drivers happen to be an equilibrium[2], then no driver will have an incentive to unilaterally switch to an alternate route. An equilibrium is called *strict* if the maximum in (1) is attained by a unique choice for every driver.

Let us now consider a repeated congestion game such that drivers adjust their routing decisions daily in response to the historical data on traffic conditions. We consider a specification of the day-to-day driver adjustment process proposed in [6]. Let $y_i(k)$ be the routing choice made by driver d_i at day $k \in \{1, 2, ...\}$, and let $y_{-i}(k)$ and $y(k)$ be defined analogously. We assume that each driver has access to the historical aggregate congestion on the road segments that are of interest to the driver. Based on such information, driver d_i can compute the average utility

$$V_i^{\bar{y}_i}(t) := \frac{1}{t} \sum_{k=1}^{t} U_i(\bar{y}_i, y_{-i}(k)) \tag{2}$$

that would have been received for the route $\bar{y}_i \in Y_i$ if $\bar{y}_i$ had been used at day t and before, assuming that the choices of all other drivers are unchanged. Before making a choice at day t, driver d_i predicts that route $\bar{y}_i \in Y_i$ would result in a utility of $V_i^{\bar{y}_i}(t-1)$. Based on this prediction, a driver d_i who believes that the choice $y_i(t-1)$ in the previous day cannot be improved upon by switching to a different route will continue to use $y_i(t-1)$ at day t, i.e., $y_i(t) = y_i(t-1)$. Otherwise, driver d_i will choose a route with maximum predicted utility with some positive probability $\alpha_i(t)$, where $\alpha_i(t)$ represents driver d_i's willingness to optimize at day t. It is clear that the drivers can completely stop adjusting their choices if and only if their choices constitute an equilibrium. However, in general, it is not clear whether the driver choices will converge to such a steady-state. As discussed earlier, the significance of convergence is that an equilibrium may be designed by a global planner as a desirable traffic pattern [10].

In the remainder of the paper, we will analyze the question of convergence by first associating the adjustment process described above with Joint Strategy Fictitious Play, a close variant of the well-known Fictitious Play process.

III. Joint Strategy Fictitious Play

We start with the well-known Fictitious Play (FP) process [1]. Consider a congestion game as introduced in the previous section with the additional assumption that the choices of all drivers are public knowledge. Define the *empirical frequency*, $q_i^{\bar{y}_i}(t)$, to be the percentage of days at which driver d_i has chosen the route $\bar{y}_i \in Y_i$ up to day t, i.e.,

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

where $y_i(k) \in Y_i$ is driver d_i's choice at day k and $I\{\cdot\}$ is the indicator function. Let us form the empirical frequency vector $q_i(t)$ by stacking up $\{q_i^{\bar{y}_i}(t)\}_{\bar{y}_i \in Y_i}$. Before making a choice at day t, driver d_i using FP predicts the utility for the route $\bar{y}_i \in Y_i$ to be

$$U_i(\bar{y}_i, q_{-i}(t)) := \sum_{y_{-i} \in Y_{-i}} U_i(\bar{y}_i, y_{-i}) \prod_{j \neq i} q_j^{y_j}(t), \tag{3}$$

where $q_{-i}(t) := \{q_1(t), ..., q_{i-1}(t), q_{i+1}(t), ..., q_n(t)\}$ and $Y_{-i} := \times_{j \neq i} Y_j$. Based on this predictions, driver d_i (randomly) selects a route at day t from the set

$$BR_i(q_{-i}(t-1)) := \{\tilde{y}_i \in Y_i : U_i(\tilde{y}_i, q_{-i}(t-1)) = \max_{y_i \in Y_i} U_i(y_i, q_{-i}(t-1))\}.$$

The set $BR_i(q_{-i}(t))$ is called driver d_i's best response to $q_{-i}(t)$.

[1]We will henceforth refer to a pure Nash equilibrium simply as an equilibrium.

[2]A pure Nash equilibrium always exist in congestion games [4].

It is known that the empirical frequencies generated by FP converge in potential games [9]. Note that FP as describe above requires each driver to observe the routing choices made by every other individual driver. Moreover, choosing a route based on the predictions (3) amounts to enumerating all possible joint choices in $\times_j Y_j$ every day for each driver. Hence, FP is not a reasonable decision making model for drivers in a large-scale congestion game.

Consider now the case in which each driver views all other drivers as a collective group, which leads us to Joint Strategy Fictitious Play (JSFP). In this case, each driver tracks the empirical frequencies of the *joint choices* of all other drivers. Let $z_{-i}^{\bar{y}_{-i}}(t)$ be the percentage of days at which drivers other then driver d_i have chosen the joint route choice profile $\bar{y}_{-i} \in Y_{-i}$ up to day t, i.e.,

$$z_{-i}^{\bar{y}_{-i}}(t) = \frac{1}{t}\sum_{k=1}^{t} I\{y_{-i}(k) = \bar{y}_{-i}\}. \qquad (4)$$

Let $z_{-i}(t)$ denote the empirical frequency vector formed by stacking up $\{z_{-i}^{\bar{y}_{-i}}(t)\}_{\bar{y}_{-i}\in Y_{-i}}$. Before choosing a route at day t, driver d_i predicts the utility for the route $\bar{y}_i \in Y_i$ as

$$U_i(\bar{y}_i, z_{-i}(t)) = \sum_{y_{-i}\in Y_{-i}} U_i(\bar{y}_i, y_{-i}) z_{-i}^{y_{-i}}(t). \qquad (5)$$

Then, at day t, driver d_i chooses a route from the set

$$BR_i(z_{-i}(t-1)) := \{\tilde{y}_i \in Y_i : U_i(\tilde{y}_i, z_{-i}(t-1)) = \max_{y_i\in Y_i} U_i(y_i, z_{-i}(t-1))\}.$$

Clearly, the computational burden of JSFP on each driver appears to be even higher than that of FP since tracking the empirical frequencies $z_{-i}(t) \in \Delta(Y_{-i})$ of the joint choices of the other drivers is more demanding for driver d_i than tracking the empirical frequencies $q_{-i}(t) \in \times_{j\neq i}\Delta(Y_j)$ of the choices of the other drivers individually, where $\Delta(Y)$ denotes the set of probability distributions on a finite set Y. However, JSFP has a connection to regret based dynamics that can be exploited to significantly reduce its computational burden on each driver. To choose a route at any day t, driver d_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 (4) into (5) yields

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

which implies

$$U_i(\bar{y}_i, z_{-i}(t)) = V_i^{\bar{y}_i}(t),$$

where $V_i^{\bar{y}_i}(t)$, introduced in (2), is the average utility driver d_i would have received if route $\bar{y}_i$ had been chosed at every day up to day t. This means that the drivers' day-to-day route adjustment process described in the previous section is in fact a JSFP process with some inertia.

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

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

which permits the JSFP dynamics to proceed without requiring each driver to track the empirical frequencies of the joint choices of the other drivers or of having to compute an expectation over the space of the joint choices of all other drivers. Each driver using JSFP merely updates the predicted utilities for each available route using the recursion above, and chooses a route every day with maximal predicted utility.

The general convergence properties of JSFP for games involving more than two players is unresolved [3]. Our simulations indicate that JSFP process is convergent in potential games with or without players using inertia, however, we are able to produce a proof only for the case where players use some inertia.

IV. Convergence of JSFP with Inertia

Here, we consider the case where drivers using JSFP are reluctant to switch to a better route. More precisely, drivers choose their routes according to the following rules:

- If the route $y_i(t-1)$ chosen by driver d_i at day $t-1$ belongs to $BR_i(z_{-i}(t-1))$, then $y_i(t) = y_i(t-1)$.
- Otherwise, driver d_i chooses a route, $y_i(t)$, at day 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 driver $\mathcal{P}_i$'s willingness to optimize at day t, $\beta_i(t) \in \Delta(Y_i)$ is a probability distribution with full support on the set $BR_i(z_{-i}(t-1))$, and $\mathbf{v}^{y_i(t-1)}$ is the vertex of $\Delta(Y_i)$ corresponding to route $y_i(t-1)$.

Thus, driver d_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 assumption on the drivers' willingness to optimize.

Assumption 4.1: There exist constants $\underline{\epsilon}$ and $\bar{\epsilon}$ such that, for all sufficiently large days t and for all drivers,

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

This assumption implies that drivers are always willing to optimize with some nonzero inertia at least asymptotically. This leads to the following convergence result.

Theorem 4.1: In any potential game, the choice profiles $y(t)$ generated by JSFP with Inertia satisfying Assumption 4.1 converge to a pure Nash equilibrium almost surely.

The proof Theorem 4.1 is omitted for the sake of brevity. It loosely follows the proof of Theorem 5.1, but with various modifications.

V. Discounting Old Information

We now extend our convergence result to the case where drivers view recent information as more important. Drivers

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

using JSFP replace true empirical frequencies with weighted empirical frequencies defined as

$$\tilde{z}_{-i}^{\bar{y}_{-i}}(t) = (1-\gamma)^t I\{y_{-i}(1) = \bar{y}_{-i}\} + \sum_{k=1}^{t} \gamma(1-\gamma)^{t-k} I\{y_{-i}(k) = \bar{y}_{-i}\},$$

where $0 < \gamma \leq 1$ is a parameter with $1-\gamma$ being the discount factor. The weighted empirical frequencies can be updated using the recursion

$$\tilde{z}_{-i}^{\bar{y}_{-i}}(t) = (1-\gamma)\tilde{z}_{-i}^{\bar{y}_{-i}}(t-1) + \gamma I\{y_{-i}(k) = \bar{y}_{-i}\}.$$

One can identify the limiting cases of the discount factor. When $\gamma = 1$ we have "Cournot" beliefs where only the most recent information matters. In the case when γ is not a constant, but rather $\gamma_t = 1/t$, all information is given equal importance as in the previous sections.

Utility prediction and route selection with exponentially discounted information are done in the same way as in the previous sections. To make a routing choice, driver d_i needs only the weighted average utility that would have been received for each route, which is defined for route $\bar{y}_i \in Y_i$ as

$$\tilde{V}_i^{\bar{y}_i}(t) := U_i(\bar{y}_i, \tilde{z}_{-i}(t)) = \sum_{y_{-i} \in Y_{-i}} U_i(\bar{y}_i, y_{-i})\tilde{z}_{-i}^{y_{-i}}(t).$$

One can easily verify that the weighted average utility $\tilde{V}_i^{\bar{y}_i}(t)$ for route $\bar{y}_i \in Y_i$ admits the recursion

$$\tilde{V}_i^{\bar{y}_i}(t) = \gamma U_i(\bar{y}_i, y_{-i}(t)) + (1-\gamma)\tilde{V}_i^{\bar{y}_i}(t-1).$$

Once again, driver d_i is not required to track the weighted empirical frequency vector $\tilde{z}_{-i}(t)$ or required to compute expectations over Y_{-i}. The following result characterizes the long-term behavior of JSFP with inertia and exponentially discounted information.

Theorem 5.1: In any potential game, the choice profiles $y(t)$ generated by JSFP with inertia satisfying Assumption 4.1 and exponentially discounted information converge to a pure Nash equilibrium almost surely.

Proof: The proof follows a similar structure as the proof of Theorem 6.2 in [12]. Fix a sufficiently large day t so that Assumption 4.1 holds true. At day $t+1$, we will have $y_i(t+1) \in BR_i(\tilde{z}_{-i}(t1))$ for every player with probability at least $\underline{\epsilon}^n$, where n is the number of players. Let $y^0 := y(t+1)$. If y^0 is an equilibrium we are done.

Otherwise, there exists a positive constant T, independent of t, such that if $y(t+1) = ... = y(t+T) = y^0$ then $BR_i(\tilde{z}_{-i}(t+T)) \subset BR_i(y^0_{-i})$ for all players. The probability of such an event is at least $(1-\bar{\epsilon})^{nT}$. Since y^0 is not an equilibrium, there must be at least one player d_i such that $y_i^0 \notin BR_i(y^0_{-i})$.

Consider now the event that, at day $t+T+1$, exactly one player switches to a different choice, i.e., $y(t+T+1) = y^1$ for some y^1 where y^0 and y^1 differs in exactly one player position. This event happens with probability at least $\underline{\epsilon}(1-\bar{\epsilon})^{n-1}$. Note that if ϕ is a potential function for the game, then $\phi(y^0) < \phi(y^1)$.

One can repeat the arguments above to construct a sequence of profiles $y^0, y^1, y^2, ..., y^M$, where M is independent of t, with the property that

$$\phi(y^0) < \phi(y^1) < ... < \phi(y^M),$$

and y^M is an equilibrium. This means that given $\{\tilde{z}_{-i}(t)\}_{i=1}^n$ at any sufficiently large t, there exist constants, $\tilde{T} > 0$ and $\tilde{\epsilon} > 0$, both of which are independent of t, such that the following event happens with probability at least $\tilde{\epsilon}$: $y(t+\tilde{T})$ is an equilibrium and $y_i(t+\tilde{T}) \in BR_i(\tilde{z}_{-i}(t+\tilde{T}-1))$ for all $i \in \{1, ..., n\}$. This implies that $y(t)$ converges to an equilibrium almost surely. ■

VI. Illustrative Example

We consider a congestion game with 100 drivers seeking to traverse from node A to node B along 10 different parallel roads as illustrated in Fig. 1. Each road is a possible route

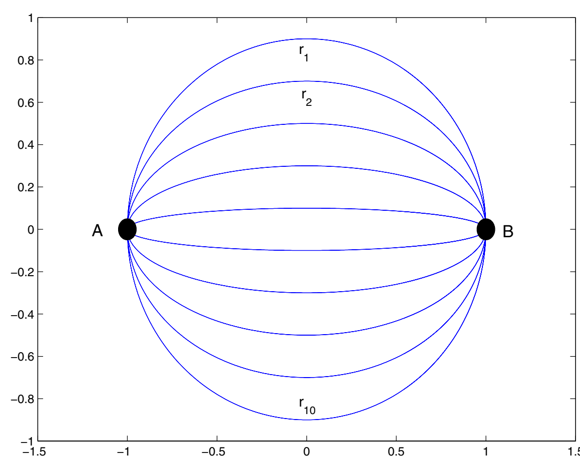

Fig. 1. Network Topology for a Congestion Game

for each driver, and has a quadratic congestion function with positive (randomly chosen) coefficients,

$$c_{r_i}(k) = \rho_1^{r_i} k^2 + \rho_2^{r_i} k + \rho_3^{r_i}, \ i = 1, ..., 10.$$

We simulated a case where drivers choose their initial routes randomly, and every day thereafter, adjust their routes using JSFP with inertia and exponentially discounted information. The parameters $\alpha_i(t)$ are chosen as 0.5 for all days and all players, whereas the parameter γ is chosen as 0.03. The number of vehicles on each road fluctuates initially and then stabilizes as illustrated in Fig. 2. Fig. 3 illustrates the evolution of the congestion cost on each road. One can observe that the congestion cost on each road converges approximately to the same value, which is consistent with a Nash equilibrium with large number of drivers. This behavior resembles an approximate "Wardrop equilibrium" [31], which represents a steady-state situation in which the congestion cost on each road is equal due to the fact that, as the number of drivers increases, the effect of an individual driver on the traffic conditions becomes negligible. Note that a driver using FP would need to track the empirical

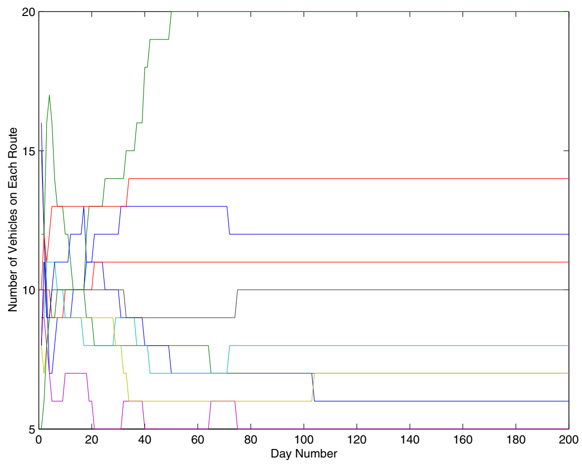

Fig. 2. Number of Vehicles on Each Route

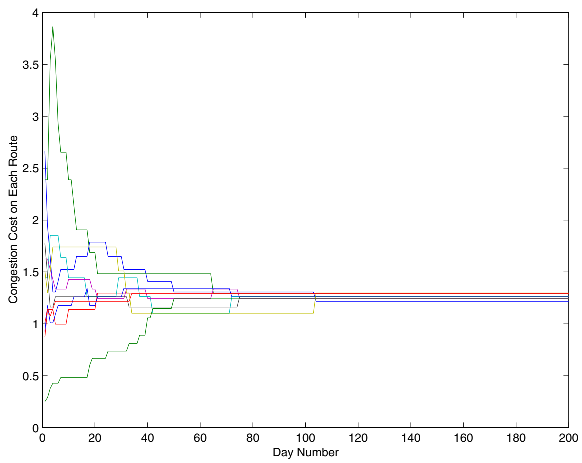

Fig. 3. Congestion Cost on each Route

frequencies of the choices of the 99 other drivers and compute an expected utility evaluated over a choice space of dimension 10^{99}.

VII. Concluding Remarks and Future Work

We have considered the long-term behavior of a large number of players in large-scale games where players are limited in both their observational and computational capabilities. The methods were motivated by and illustrated on a transportation congestion game, in which a large number of vehicles make daily routing decisions to optimize their own objectives in response to the aggregate congestion on each road of interest. In particular, we analyzed a version of JSFP and showed that it accommodates inherent player limitations in information gathering and processing. Furthermore, we showed that JSFP has guaranteed convergence to a pure Nash equilibrium in congestion games, or equivalently in finite potential games [9], when players use some inertia and either with or without exponential discounting of the historical data. An important continuation of this research would be the case where players observe only the actual utilities they receive.

References

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

[2] D. Fudenberg and D. Kreps, "Learning mixed equilibria," *Games and Economic Behavior*, vol. 5, no. 3, pp. 320–367, 1993.

[3] D. Monderer and A. Sela, "Fictitious play and no-cycling conditions," 1997, http://www.sfb504.uni-mannheim.de/publications/dp97-12.pdf.

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

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

[6] M. Ben-Akiva, A. de Palma, and I. Kaysi, "Dynamic network models and driver information systems," *Transportation Research A*, vol. 25A, pp. 251–266, 1991.

[7] M. Jha, S. Madanat, and S. Peeta, "Perception updating and day-to-day travel choice dynamics in traffic networks with information provision," *Transportation Research C*, vol. 6, pp. 189–212, 1998.

[8] E. Cascetta and G. E. Cantarella, "A day-to-day and within-day dynamic stochastic assignment model," *Transportation Research A*, vol. 25A, pp. 277–291, 1991.

[9] D. Monderer and L. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.

[10] W. Sandholm, "Evolutionary implementation and congestion pricing," *Review of Economic Studies*, vol. 69, no. 3, pp. 667–689, 2002.

[11] H. P. Young, *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press, 1998.

[12] ——, *Strategic Learning And Its Limits*. Oxford University Press, 2005.

[13] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[14] J. Weibull, *Evolutionary Game Theory*. Cambridge, MA: MIT Press, 1995.

[15] S. Hart, "Adaptive heuristics," *Econometrica*, vol. **73**, no. 5, pp. 1401–1430, 2005.

[16] S. Hart and A. Mas-Colell, "Uncoupled dynamics do not lead to Nash equilibrium," *American Economic Review*, vol. **93**, no. 5, pp. 1830–1836, 2003.

[17] J. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Transactions on Automatic Control*, vol. **50**, no. 3, pp. 312–327, 2005.

[18] G. Arslan and J. Shamma, "Distributed convergence to Nash equilibria with local utility measurements," in *43rd IEEE Conference on Decision and Control*, 2004, pp. 1538–1543.

[19] T. Lambert, M. Epelman, and R. Smith, "A fictitious play approach to large-scale optimization," *Operations Research*, 2005, forthcoming.

[20] S. Takahashi and T. Yamamori, "The global convergence of learning processes in supermodular games," 2003.

[21] H. P. Young, "The evolution of conventions," *Econometrica*, vol. 61, no. 1, pp. 57–84, 1993.

[22] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, 2000.

[23] ——, "Regret based continuous-time dynamics," *Games and Economic Behavior*, vol. **45**, pp. 375–394, 2003.

[24] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MA: MIT Press, 1998.

[25] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[26] D. Leslie and E. Collins, "Convergent multiple-timescales reinforcement learning algorithms in normal form games," *Annals of Applied Probability*, vol. 13, pp. 1231–1251, 2003.

[27] ——, "Individual q-learning in normal form games," *SIAM Journal on Control and Optimization*, submitted. [Online]. Available: http://www.stats.ox.ac.uk/~dleslie/papers/LeslieCollinsSICON04.pdf

[28] ——, "Generalised weakened fictitious play," submitted. [Online]. Available: http://www.stats.ox.ac.uk/~dleslie/papers/LeslieCollinsWeakenedFP.pdf

[29] S. Huck and R. Sarin, "Players with limited memory," *Contributions to Theoretical Economics*, vol. 4, no. 1, 2004.

[30] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, PA: SIAM, 1999.

[31] J. Wardrop, "Some theoretical aspects of road traffic research," in *Proceedings of the Institute of Civil Engineers*, vol. I, pt. II, London, Dec. 1952, pp. 325–378.