Joint Strategy Fictitious Play With Inertia for Potential Games

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

***Abstract*—We consider 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. This disqualifies some of the well known decision making models such as "Fictitious Play" (FP), in which each player must monitor the individual actions of every other player and must optimize over a high dimensional probability space. We will show that Joint Strategy Fictitious Play (JSFP), a close variant of FP, alleviates both the informational and computational burden of FP. Furthermore, we introduce JSFP with inertia, i.e., a probabilistic reluctance to change strategies, and establish the convergence to a pure Nash equilibrium in all generalized ordinal potential games in both cases of averaged or exponentially discounted historical data. We illustrate JSFP with inertia on the specific class of congestion games, a subset of generalized ordinal potential games. In particular, we illustrate the main results on a distributed traffic routing problem and derive tolling procedures that can lead to optimized total traffic congestion.**

***Index Terms*—Fictitious play (FP), joint strategy fictitious play (JSFP).**

I. Introduction

We consider "large-scale" repeated games involving a large number of players, each of whom selects a strategy from a possibly large strategy set. A player's reward, or utility, depends on the actions taken by all players. The game is repeated over multiple stages, and this allows players to adapt their strategies in response to the available information gathered over prior stages. This setup falls under the general subject of "learning in games" [2], [3], and there are a variety of algorithms and accompanying analysis that examine the long term behavior of these algorithms.

In 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. This limits the feasibility of different learning algorithms. For example, the well-studied algorithm "Fictitious Play" (FP) requires individual players to individually monitor the actions of other players and to optimize their strategies according to a probability distribution function over the joint actions of other players. Clearly, such information gathering and processing is not feasible in a large-scale game.

The main objective of this paper [1] is to study a variant of FP called Joint Strategy Fictitious Play (JSFP) [2], [4], [5]. We will argue that JSFP is a plausible decision making model for certain large-scale games. We will introduce a modification of JSFP to include inertia, in which there is a probabilistic reluctance of any player to change strategies. We will establish that JSFP with inertia converges to a pure Nash equilibrium for a class of games known as generalized ordinal potential games, which includes so-called congestion games as a special case [6].

Our motivating example for a large-scale congestion game is distributed traffic routing [7], 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. A more realistic measurement on the information tracked and processed by an individual driver is the daily aggregate congestion on the roads that are of interest to that driver [8]. It turns out that JSFP accommodates such information aggregation.

We will now review some of the well known decision making models and discuss their limitations in large-scale games. See the monographs [2], [3], [9]–[11] and survey article [12] for a more comprehensive review.

The well known FP algorithm requires that each player views all other players as independent decision makers [2]. 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 [13]–[15].

The paper [16] introduces a version of FP, called "sampled FP", that seeks to avoid computing an expected utility based on

Manuscript received December 07, 2006. Current version published February 11, 2009. This work was supported by NSF Grants CMS-0339228, ECS-0501394, and ECCS-0547692, and ARO Grant W911NF-04-1-0316. This paper appeared in part at the 44th IEEE Conference on Decision and Control, 2005. Recommended by Associate Editor F. Bullo.

J. R. Marden is with the Social and Information Sciences Laboratory, California Institute of Technology, Pasadena, CA 91107 USA (e-mail: marden@caltech.edu).

G. Arslan is with the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, HI 96822 USA (e-mail: gurdal@hawaii.edu).

J. S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA (e-mail: shamma@gatech.edu.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2008.2010885

the empirical frequencies, because for large scale games, this expected utility computation can be prohibitively demanding. 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. A player then 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. Sampled FP is proved to be convergent in identical interest games, but the number of samples needed to guarantee convergence grows unboundedly.

There are convergent learning algorithms for a large class of coordination games called "weakly acyclic" games [9]. In adaptive play [17] players have finite recall and respond to the recent history of other players. Adaptive play requires each player to track the individual behavior of all other players for recall window lengths 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 the JSFP discussed herein and the regret matching algorithm [18]. 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. In JSFP, a player chooses an action by myopically maximizing the anticipated utility based on past observations, which is effectively equivalent to regret modulo a bias term. 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 [19]). There are finite memory versions of the regret matching algorithm and various generalizations [3], 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.

There are also payoff based dynamics, where each player observes only the actual utilities received and uses a Reinforcement Learning (RL) algorithm [20], [21] to make future choices. Convergence of player choices when all players use an RL-like algorithm is proved for identical interest games [22]–[24] assuming that learning takes place at multiple time scales. Finally, the payoff based dynamics with finite-memory presented in [25] leads to a Pareto-optimal outcome in generic common interest games.

Regarding the distributed routing setting of Section IV, there are papers that analyze different routing strategies in congestion games with "infinitesimal" players, i.e., a continuum of players as opposed to a large, but finite, number of players. References [26]–[28] analyze the convergence properties of a class of routing strategies that is a variation of the replicator dynamics in congestion games, also referred to as symmetric games, under a variety of settings. Reference [29] analyzes the convergence properties of no-regret algorithms in such congestion games and also considers congestion games with discrete players, as considered in this paper, but the results hold only for a highly structured symmetric game.

The remainder of the paper is organized as follows. Section II, sets up JSFP and goes on to establish convergence to a pure Nash equilibrium for JSFP with inertia in all generalized ordinal potential games. Section III presents a fading memory variant of JSFP, and likewise establishes convergence to a pure Nash equilibrium. Section IV presents an illustrative example for traffic congestion games. Section IV goes on to illustrate the use of tolls to achieve a socially optimal equilibrium and derives conditions for this equilibrium to be unique. Finally, Section V presents some concluding remarks.

II. Joint Strategy Fictitious Play With Inertia

A. Setup

Consider a finite game with n-player set $\mathcal{P} := \{\mathcal{P}_1, \ldots, \mathcal{P}_n\}$ 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 $Y = Y_1 \times \ldots \times Y_n$.

For $y = (y_1, y_2, \ldots, y_n) \in Y$, let y_{-i} denote the profile of player actions *other than* player $\mathcal{P}_i$, 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 actions as (y_i, y_{-i}). Similarly, we may write $U_i(y)$ as $U_i(y_i, y_{-i})$.

A profile $y^* \in Y$ of actions is called a *pure Nash equilibrium*[1] if, for all players $\mathcal{P}_i \in \mathcal{P}$

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

We will consider the class of games known as "generalized ordinal potential games", defined as follows.

Definition 2.1 (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 **potential game** if, for some potential function $\phi : Y_1 \times \ldots \times Y_n \to \mathbb{R}$

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

for every player, for every $y_{-i} \in \times_{j \neq i} Y_j$ and for every $y_i', y_i'' \in Y_i$. It is a **generalized ordinal potential game** if, for some potential function $\phi : Y_1 \times \ldots \times Y_n \to \mathbb{R}$

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

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

In a *repeated* version of this setup, at every stage $t \in \{0, 1, 2, \ldots\}$, each player, $\mathcal{P}_i$, selects an action $y_i(t) \in Y_i$. This selection is a function of the information available to player $\mathcal{P}_i$ up to stage t. Both the action selection function and

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

the available information depend on the underlying learning process.

B. Fictitious Play

We start with the well known Fictitious Play (FP) process [2].

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.,

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

where $y_i(k) \in Y_i$ is player $\mathcal{P}_i$'s action at time k and $I\{\cdot\}$ is the indicator function. Now define the empirical frequency vector for player $\mathcal{P}_i$ as

$$q_i(t) := \begin{pmatrix} q_i^{\bar{y}_1} \\ \vdots \\ q_i^{\bar{y}_{|Y_i|}} \end{pmatrix}$$

where $|Y_i|$ is the cardinality of the action set 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

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

where $q_{-i}(t) := \{q_1(t), \ldots, q_{i-1}(t), q_{i+1}(t), \ldots, q_n(t)\}$ 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 [30].

Note that FP as described 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 $\times_j Y_j$ at every stage for each player. Hence, FP is computationally prohibitive as a decision making model in large-scale games.

C. JSFP

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 $z^{\bar{y}}(t)$ be the percentage of stages at which all players chose the joint action profile $\bar{y} \in Y$ up to time $t-1$, i.e.,

$$z^{\bar{y}}(t) := \frac{1}{t}\sum_{\tau=0}^{t-1} I\left\{y(\tau) = \bar{y}\right\}. \tag{3}$$

Let $z(t)$ denote the empirical frequency vector formed by the components $\{z^{\bar{y}}(t)\}_{\bar{y}\in Y}$. Note that the dimension of $z(t)$ is the cardinality $|Y|$.

Similarly, 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.,

$$z_{-i}^{\bar{y}_{-i}}(t) := \frac{1}{t}\sum_{\tau=0}^{t-1} I\left\{y_{-i}(\tau) = \bar{y}_{-i}\right\}, \tag{4}$$

which, given $z(t)$, can also be expressed as

$$z_{-i}^{\bar{y}_{-i}}(t) = \sum_{y_i \in Y_i} z^{(y_i, \bar{y}_{-i})}(t).$$

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[2]

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

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

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

Note that the utility as expressed in (5) is linear in $z_{-i}(t)$.

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 (4) into (5) results in

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

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. Let $\bar{U}_i^{\bar{y}_i}(t) := U_i(\bar{y}_i, z_{-i}(t))$. This average utility, $\bar{U}_i^{\bar{y}_i}(t)$, admits the following simple recursion

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

[2] 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 (2) and (5), respectively.

The important implication is that JSFP dynamics can be implemented *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. Rather, each player using JSFP merely updates the predicted utilities for each available action using the recursion above, and chooses an action each stage with maximal predicted utility.

An interesting feature of JSFP is that each strict Nash equilibrium has an "absorption" property as summarized in Proposition 2.1.

Proposition 2.1: In any finite n-person game, if at any time $t > 0$, the joint action $y(t)$ generated by a JSFP process is a strict Nash equilibrium, then $y(t+\tau) = y(t)$ for all $\tau > 0$.

Proof: For each player $\mathcal{P}_i \in \mathcal{P}$ and for all actions $y_i \in Y_i$,

$$U_i\left(y_i(t), z_{-i}(t)\right) \geq U_i\left(y_i, z_{-i}(t)\right).$$

Since $y(t)$ is a strict Nash equilibrium, we know that for all actions $y_i \in Y_i \setminus y_i(t)$

$$U_i\left(y_i(t), y_{-i}(t)\right) > U_i\left(y_i, y_{-i}(t)\right).$$

By writing $z_{-i}(t+1)$ in terms of $z_{-i}(t)$ and $y_{-i}(t)$

$$U_i(y_i(t), z_{-i}(t+1)) = \frac{t}{t+1} U_i(y_i(t), z_{-i}(t)) + \frac{1}{t+1} U_i(y_i(t), y_{-i}(t)).$$

Therefore, $y_i(t)$ is the only best response to $z_{-i}(t+1)$, i.e., for all $y_i \in Y_i \setminus y_i(t)$

$$U_i\left(y_i(t), z_{-i}(t+1)\right) > U_i\left(y_i, z_{-i}(t+1)\right).$$

□

A strict Nash equilibrium need *not* possess this absorption property in general for standard FP when there are more than two players.[3]

The convergence properties, even for potential games, of JSFP in the case of more than two players is unresolved.[4] We will establish convergence of JSFP in the case where players use some sort of inertia, i.e., players are reluctant to switch to a better action.

D. JSFP 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 probability distribution with full support on the action $y_i(t-1)$, i.e.,

$$\mathbf{v}^{y_i(t-1)} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where the "1" occurs in the coordinate of $\Delta(Y_i)$ associated with $y_i(t-1)$.

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 $\mathcal{P}_i \in \mathcal{P}$

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

This assumption implies that players are always willing to optimize with some nonzero inertia.[5]

The following result shows a similar absorption property of pure Nash equilibria in a JSFP with inertia process.

Proposition 2.2: In any finite n-person game, if at any time $t > 0$ the joint action $y(t)$ generated by a JSFP with inertia process is 1) a pure Nash equilibrium and 2) the action $y_i(t) \in BR_i(z_{-i}(t))$ for all players $\mathcal{P}_i \in \mathcal{P}$, then $y(t+\tau) = y(t)$ for all $\tau > 0$.

We will omit the proof of Proposition 2.2 as it follows very closely to the proof of Proposition 2.1.

E. Convergence to Nash Equilibrium

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

We will assume that no player is indifferent between distinct strategies.[6]

Assumption 2.2: Player utilities satisfy the following: for all players $\mathcal{P}_i \in \mathcal{P}$, actions $y_i^1, y_i^2 \in Y_i$, $y_i^1 \neq y_i^2$, and joint actions $y_{-i} \in Y_{-i}$

$$U_i\left(y_i^1, y_{-i}\right) \neq U_i\left(y_i^2, y_{-i}\right). \tag{6}$$

Theorem 2.1: In any finite generalized ordinal potential game in which no player is indifferent between distinct strategies as in

[3] To see this, consider the following 3 player identical interest game. For all $\mathcal{P}_i \in \mathcal{P}$, let $Y_i = \{a, b\}$. Let the utility be defined as follows: $U(a,b,a) = U(b,a,a) = 1$, $U(a,a,a) = U(b,b,a) = 0$, $U(a,a,b) = U(b,b,b) = 1$, $U(a,b,b) = -1$, $U(b,a,b) = -100$. Suppose the first action played is $y(1) = \{a,a,a\}$. In the FP process each player will seek to deviate in the ensuing stage, $y(2) = \{b,b,b\}$. The joint action $\{b,b,b\}$ is a strict Nash equilibrium. One can easily verify that the ensuing action in a FP process will be $y(3) = \{a,b,a\}$. Therefore, a strict Nash equilibrium is not absorbing in the FP process with more than 2 players.

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

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

[6] One could alternatively assume that all pure equilibria are strict.

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.

We provide a complete proof of Theorem 2.1 in the Appendix. We encourage the reader to first review the proof of fading memory JSFP with inertia in Theorem 3.1 of the following section.

F. Relationship Between Regret Matching and JSFP

It turns out that JSFP is strongly related to the learning algorithm regret matching, from [18], in which players choose their actions based on their *regret* for not choosing particular actions in the past steps.

Define the average regret of player $\mathcal{P}_i$ for an action $y_i \in Y_i$ at time t as

$$R_i^{y_i}(t) := \frac{1}{t}\sum_{\tau=0}^{t-1}\left(U_i\left(y_i, y_{-i}(\tau)\right) - U_i\left(y(\tau)\right)\right). \tag{7}$$

In other words, player $\mathcal{P}_i$'s average regret for $y_i \in Y_i$ would represent the average improvement in his utility if he had chosen $y_i \in Y_i$ in all past steps and all other players' actions had remained unaltered. Notice that the average regret in (7) can also be expressed in terms of empirical frequencies, i.e.,

$$R_i^{y_i}(t) = U_i\left(y_i, z_{-i}(t)\right) - U_i\left(z(t)\right)$$

where

$$U_i\left(z(t)\right) := \sum_{y\in Y} U_i(y) z^y(t) = \frac{1}{t}\sum_{\tau=0}^{t-1} U_i\left(y(\tau)\right).$$

In regret matching, once player $\mathcal{P}_i$ computes his average regret for each action $y_i \in Y_i$, he chooses an action $y_i(t)$, $t > 0$, according to the probability distribution $p_i(t)$ defined as

$$p_i^{y_i}(t) = \mathbf{Pr}\left[y_i(t) = y_i\right] = \frac{\left[R_i^{y_i}(t)\right]^+}{\sum_{\tilde{y}_i\in Y_i}\left[R_i^{\tilde{y}_i}(t)\right]^+}$$

for any $y_i \in Y_i$, provided that the denominator above is positive; otherwise, $p_i(t)$ is the uniform distribution over Y_i. Roughly speaking, a player using regret matching chooses a particular action at any step with probability proportional to the average regret for not choosing that particular action in the past steps. This is in contrast to JSFP, where each player would only select the action that yielded the highest regret.

If all players use regret matching, then the empirical frequency $z(t)$ of the joint actions converges almost surely to the set of coarse correlated equilibria, a generalization of Nash equilibria, in any game [18]. We prove that if all players use JSFP with inertia, then the action profile converges almost surely to a pure Nash equilibrium, albeit in the special glass of generalized ordinal potential games. The convergence properties of regret matching (with or without inertia) in potential games remains an open question.

III. Fading Memory JSFP With Inertia

We now analyze the case where players view recent information as more important. In fading memory JSFP with inertia, players replace true empirical frequencies with weighted empirical frequencies defined by the recursion

$$\begin{aligned}\tilde{z}_{-i}^{\bar{y}_{-i}}(0) &:= I\left\{y_{-i}(0) = \bar{y}_{-i}\right\}\\ \tilde{z}_{-i}^{\bar{y}_{-i}}(t) &:= (1-\rho)\tilde{z}_{-i}^{\bar{y}_{-i}}(t-1) + \rho I\left\{y_{-i}(t-1) = \bar{y}_{-i}\right\}\end{aligned}$$

for all times $t \geq 1$ where $0 < \rho \leq 1$ is a parameter with $1-\rho$ being the discount factor. Let $\tilde{z}_{-i}(t)$ denote the weighted empirical frequency vector formed by the components $\{\tilde{z}_{-i}^{\bar{y}_{-i}}(t)\}_{\bar{y}_{-i}\in Y_{-i}}$. Note that the dimension of $\tilde{z}_{-i}(t)$ is the cardinality $|Y_{-i}|$.

One can identify the limiting cases of the discount factor. When $\rho = 1$ we have "Cournot" beliefs, where only the most recent information matters. In the case when ρ is not a constant, but rather $\rho(t) = 1/t$, all past information is given equal importance as analyzed in Section II.

Utility prediction and action selection with fading memory are done in the same way as in Section II, and in particular, in accordance with rules JSFP-1 and JSFP-2. To make a decision, player $\mathcal{P}_i$ needs only the weighted average utility that would have been received for each action, which is defined for action $\bar{y}_i \in Y_i$ as

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

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

$$\tilde{U}_i^{\bar{y}_i}(t) = \rho U_i\left(\bar{y}_i, y_{-i}(t-1)\right) + (1-\rho)\tilde{U}_i^{\bar{y}_i}(t-1).$$

Once again, player $\mathcal{P}_i$ is not required to track the weighted empirical frequency vector $\tilde{z}_{-i}(t)$ or required to compute expectations over Y_{-i}.

As before, pure Nash equilibria have an absorption property under fading memory JSFP with inertia.

Proposition 3.1: In any finite n-person game, if at any time $t > 0$ the joint action $y(t)$ generated by a fading memory JSFP with inertia process is 1) a pure Nash equilibrium and 2) the action $y_i(t) \in BR_i(\tilde{z}_{-i}(t))$ for all players $\mathcal{P}_i \in \mathcal{P}$, then $y(t+\tilde{t}) = y(t)$ for all $\tilde{t} > 0$.

We will omit the proof of Proposition 3.1 as it follows very closely to the proof of Proposition 2.1.

The following theorem establishes convergence to Nash equilibrium for fading memory JSFP with inertia.

Theorem 3.1: 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 a fading memory JSFP with inertia process satisfying Assumption 2.1 converge to a pure Nash equilibrium almost surely.

Proof: The proof follows a similar structure to the proof of Theorem 6.2 in [3]. At time t, let $y^0 := y(t)$. There exists a positive constant T, independent of t, such that if the current action y^0 is repeated T consecutive stages, i.e. $y(t) = \ldots = y(t+T-1) = y^0$, then $BR_i(\tilde{z}_{-i}(t+T)) = BR_i(y_{-i}^0)$

for all players.[7] The probability of such an event is at least $(1-\overline{\varepsilon})^{n(T-1)}$, where n is the number of players. If the joint action y^0 is an equilibrium, then by Proposition 3.1 we are done. Otherwise, there must be at least one player $\mathcal{P}_{i(1)} \in \mathcal{P}$ such that $y^0_{i(1)} \notin BR_{i(1)}(y^0_{-i(1)})$ and hence $y^0_{i(1)} \notin BR_{i(1)}(\tilde{z}_{-i(1)}(t+T))$.

Consider now the event that, at time $t+T$, exactly one player switches to a different action, i.e., $y^1 := y(t+T) = (y^*_{i(1)}, y^0_{-i(1)})$ for some player $\mathcal{P}_{i(1)} \in \mathcal{P}$ where $U_{i(1)}(y^1) > U_{i(1)}(y^0)$. This event happens with probability at least $\underline{\varepsilon}(1-\overline{\varepsilon})^{n-1}$. Note that if $\phi(\cdot)$ is a generalized ordinal potential function for the game, then $\phi(y^0) < \phi(y^1)$.

Continuing along the same lines, if the current action y^1 is repeated T consecutive stages, i.e. $y(t+T) = \ldots = y(t+2T-1) = y^1$, then $BR_i(\tilde{z}_{-i}(t+2T)) = BR_i(y^1_{-i})$ for all players. The probability of such an event is at least $(1-\overline{\varepsilon})^{n(T-1)}$. If the joint action y^1 is an equilibrium, then by Proposition 3.1, we are done. Otherwise, there must be at least one player $\mathcal{P}_{i(2)} \in \mathcal{P}$ such that $y^1_{i(2)} \notin BR_{i(2)}(y^1_{-i(2)})$ and hence $y^1_{i(2)} \notin BR_{i(2)}(\tilde{z}_{-i(2)}(t+2T))$.

One can repeat the arguments above to construct a sequence of profiles $y^0, y^1, y^2, \ldots, y^m$, where $y^k = (y^*_{i(k)}, y^{k-1}_{-i(k)})$ for all $k \geq 1$, with the property that

$$\phi(y^0) < \phi(y^1) < \ldots < \phi(y^m)$$

and y^m is an equilibrium. This means that given $\{\tilde{z}_{-i}(t)\}_{i=1}^n$, there exist constants

$$\tilde{T} = (|Y|+1)T > 0$$
$$\tilde{\varepsilon} = \left(\underline{\varepsilon}(1-\overline{\varepsilon})^{n-1}\right)^{|Y|} \left((1-\overline{\varepsilon})^{n(T-1)}\right)^{|Y|+1} > 0$$

both of which are independent of t, such that the following event happens with probability at least $\tilde{\varepsilon} : y(t+\tilde{T})$ is an equilibrium and $y_i(t+\tilde{T}) \in BR_i(\tilde{z}_{-i}(t+\tilde{T}))$ for all players $\mathcal{P}_i \in \mathcal{P}$. This implies that $y(t)$ converges to a pure equilibrium almost surely. □

IV. Congestion Games and Distributed Traffic Routing

In this section, we illustrate the main results on congestion games, which are a special case of the generalized ordinal potential games addressed in Theorems 2.1 and 3.1. We first recall the definition of player utilities that constitute a congestion game. We illustrate these results on a simulation of distributed traffic routing. We go on to discuss how to modify player utilities in distributed traffic routing to allow a centralized planner to achieve a desired collective objective through distributed learning.

A. Congestion Games

Congestion games are a specific class of games in which player utility functions have a special structure.

In order to define a congestion game, we must specify the action set, Y_i, and utility function, $U_i(\cdot)$, of each player. Towards this end, let $\mathcal{R}$ denote a finite set of "resources". For each resource $r \in \mathcal{R}$, there is an associated "congestion function"

$$c_r : \{0, 1, 2, \ldots\} \to \mathbb{R}$$

that reflects the cost of using the resource as a function of the number of players using that resource.

The action set, Y_i, of each player, $\mathcal{P}_i$, is defined as the set of resources available to player $\mathcal{P}_i$, i.e.,

$$Y_i \subset 2^{\mathcal{R}}$$

where $2^{\mathcal{R}}$ denotes the set of subsets of $\mathcal{R}$. Accordingly, an action, $y_i \in Y_i$, reflects a selection of (multiple) resources, $y_i \subset \mathcal{R}$. A player is "using" resource r if $r \in y_i$. For an action profile $y \in Y_1 \times \ldots \times Y_n$, let $\sigma_r(y)$ denote the total number of players using resource r, i.e., $|\{i : r \in y_i\}|$. In a congestion game, the utility of player $\mathcal{P}_i$ using resources indicated by y_i depends only on the total number of players using the same resources. More precisely, the utility of player $\mathcal{P}_i$ is defined as

$$U_i(y) = -\sum_{r \in y_i} c_r\left(\sigma_r(y)\right). \quad (8)$$

The negative sign stems from $c_r(\cdot)$ reflecting the cost of using a resource and $U_i(\cdot)$ reflecting a utility or reward function. Any congestion game with utility functions as in (8) is a potential game [6].[8]

A congestion game can be generalized further by allowing player utilities to include player specific attributes [31]. For example, each player may have a personal preference over resources, in which case player utilities take the form

$$U_i(y) = -\sum_{r \in y_i} \left(c_r\left(\sigma_r(y)\right) + f_{r,i}\right)$$

where $-f_{r,i}$ is the fixed utility player $\mathcal{P}_i$ receives for using resource r. Congestion games of this form are also potential games [31].

B. Distributed Traffic Routing

We consider a simple scenario with 100 players (drivers) seeking to traverse from node A to node B along 10 different parallel roads as illustrated in Fig. 1. Each driver can select any road as a possible route. In terms of congestion games, the set of resources is the set of roads, $\mathcal{R}$, and each player can select one road, i.e., $Y_i = \mathcal{R}$.

Each road has a quadratic cost function with positive (randomly chosen) coefficients,

$$c_{r_i}(k) = a_i k^2 + b_i k + c_i, \quad i = 1, \ldots, 10$$

[7] To see this, notice that at time $t+T$, the weighted empirical frequencies are equal to

$$\tilde{z}_{-i}(t+T) = \left(1-(1-\rho)^T\right)\mathbf{v}^{y^0_{-i}} + (1-\rho)^T \tilde{z}_{-i}(t).$$

Therefore, when T is sufficiently large, the best response set, $BR_i(\tilde{z}_{-i}(t+T))$, does not depend on the old weighted empirical frequencies, $\tilde{z}_{-i}(t)$. Furthermore, note that this sufficiently large time T is independent of t. Since no player is indifferent between distinct strategies, the best response to the current action profile, $BR_i(y^0_{-i})$, is a singleton.

[8] In fact, every congestion game is a potential game and every finite potential game is isomorphic to a congestion game [30].

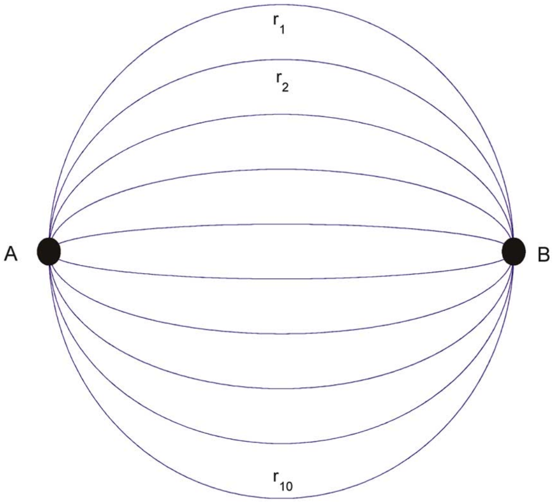

Fig. 1. Network topology for a congestion game.

where k represent the number of vehicles on that particular road. The actual coefficients are unimportant as we are just using this example as an opportunity to illustrate the convergence properties of the algorithm fading memory JSFP with inertia. This cost function may represent the delay incurred by a driver as a function of the number of other drivers sharing the same road.

We simulated a case where drivers choose their initial routes randomly, and every day thereafter, adjusted their routes using fading memory JSFP with inertia. The parameters $\alpha_i(t)$ are chosen as 0.5 for all days and all players, and the fading memory parameter ρ is chosen as 0.03. The number of vehicles on each road fluctuates initially and then stabilizes at a Nash equilibrium 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" [32], 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 FP could not be implemented even on this very simple congestion game. A driver using FP would need to track the empirical frequencies of the choices of the 99 other drivers and compute an expected utility evaluated over a probability space of dimension 10^{99}.

It turns out that both JSFP and fading memory JSFP are strongly connected to actual driver behavioral models. Consider the driver adjustment process considered in [8] which is illustrated in Fig. 4. The adjustment process highlighted is precisely JSFP with Inertia.

C. Incorporating Tolls to Minimize the Total Congestion

It is well known that a Nash equilibrium may not minimize the total congestion experienced by all drivers [33]. In this section, we show how a global planner can minimize the total congestion by implementing tolls on the network. The results are applicable to general congestion games, but we present the approach in the language of distributed traffic routing.

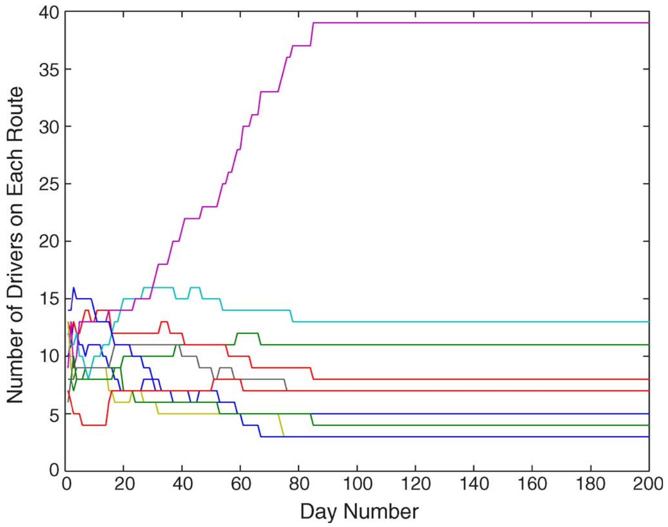

Fig. 2. Evolution of number of vehicles on each route.

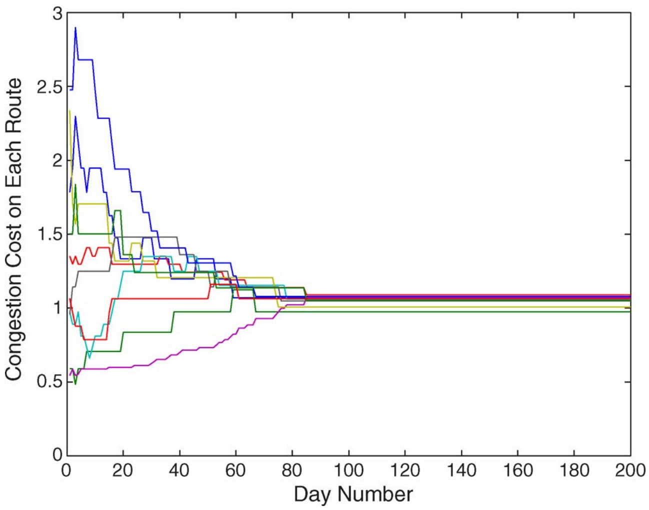

Fig. 3. Evolution of congestion cost on each route.

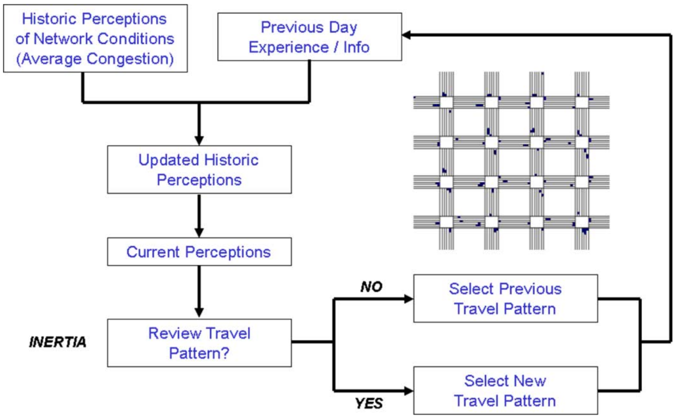
Historic Perceptions of Network Conditions (Average Congestion)
Previous Day Experience / Info
Updated Historic Perceptions
Current Perceptions
INERTIA
Review Travel Pattern?
NO
Select Previous Travel Pattern
YES
Select New Travel Pattern

Fig. 4. Example of a driver adjustment process.

The total congestion experienced by all drivers on the network is

$$T_c(y) := \sum_{r \in \mathcal{R}} \sigma_r(y) c_r\left(\sigma_r(y)\right).$$

Define a new congestion game where each driver's utility takes the form

$$U_i(y) = -\sum_{r\in y_i} \left(c_r\left(\sigma_r(y)\right) + t_r\left(\sigma_r(y)\right)\right)$$

where $t_r(\cdot)$ is the toll imposed on road r which is a function of the number of users of road r.

The following proposition outlines how to incorporate tolls so that the minimum congestion solution is a Nash equilibrium. The approach is similar to the taxation approaches for nonatomic congestion games proposed in [31], [34].

Proposition 4.1: Consider a congestion game of any network topology. If the imposed tolls are set as

$$t_r(k) = (k-1)\left[c_r(k) - c_r(k-1)\right], \quad \forall k \geq 1$$

then the total negative congestion experienced by all drivers, $\phi_c(y) := -T_c(y)$, is a potential function for the congestion game with tolls.

Proof: Let $y^1 = \{y_i^1, y_{-i}\}$ and $y^2 = \{y_i^2, y_{-i}\}$. We will use the shorthand notation $\sigma_r^{y^1}$ to represent $\sigma_r(y^1)$. The change in utility incurred by the i^{th} driver in changing from route y_i^2 to route y_i^1 is

$$\begin{aligned}
U_i(y^1)-U_i(y^2) &\\
=-\sum_{r\in y_i^1}&(c_r(\sigma_r^{y^1})+t_r(\sigma_r^{y^1}))+\sum_{r\in y_i^2}(c_r(\sigma_r^{y^2})+t_r(\sigma_r^{y^2}))\\
=-\sum_{r\in y_i^1\setminus y_i^2}&(c_r(\sigma_r^{y^1})+t_r(\sigma_r^{y^1}))+\sum_{r\in y_i^2\setminus y_i^1}(c_r(\sigma_r^{y^2})+t_r(\sigma_r^{y^2})).
\end{aligned}$$

The change in the total negative congestion from the joint action y^2 to y^1 is

$$\phi_c(y^1)-\phi_c(y^2)=-\sum_{r\in\left(y_i^1\cup y_i^2\right)}\left(\sigma_r^{y^1}c_r\left(\sigma_r^{y^1}\right)-\sigma_r^{y^2}c_r\left(\sigma_r^{y^2}\right)\right).$$

Since

$$\sum_{r\in\left(y_i^1\cap y_i^2\right)}\left(\sigma_r^{y^1}c_r\left(\sigma_r^{y^1}\right)-\sigma_r^{y^2}c_r\left(\sigma_r^{y^2}\right)\right)=0$$

the change in the total negative congestion is

$$\begin{aligned}
\phi_c(y^1)-\phi_c(y^2) = &-\sum_{r\in y_i^1\setminus y_i^2}\left(\sigma_r^{y^1}c_r\left(\sigma_r^{y^1}\right)-\sigma_r^{y^2}c_r\left(\sigma_r^{y^2}\right)\right)\\
&-\sum_{r\in y_i^2\setminus y_i^1}\left(\sigma_r^{y^1}c_r\left(\sigma_r^{y^1}\right)-\sigma_r^{y^2}c_r\left(\sigma_r^{y^2}\right)\right).
\end{aligned}$$

Expanding the first term, we obtain

$$\begin{aligned}
&\sum_{r\in y_i^1\setminus y_i^2}\left(\sigma_r^{y^1}c_r\left(\sigma_r^{y^1}\right)-\sigma_r^{y^2}c_r\left(\sigma_r^{y^2}\right)\right)\\
&\quad=\sum_{r\in y_i^1\setminus y_i^2}\left(\sigma_r^{y^1}c_r\left(\sigma_r^{y^1}\right)-\left(\sigma_r^{y^1}-1\right)c_r\left(\sigma_r^{y^1}-1\right)\right)\\
&\quad=\sum_{r\in y_i^1\setminus y_i^2}\left(c_r\left(\sigma_r^{y^1}\right)+t_r\left(\sigma_r^{y^1}\right)\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\phi_c(y^1)-\phi_c(y^2)\\
&=-\sum_{r\in y_i^1\setminus y_i^2}(c_r(\sigma_r^{y^1})+t_r(\sigma_r^{y^1}))+\sum_{r\in y_i^2\setminus y_i^1}(c_r(\sigma_r^{y^2})+t_r(\sigma_r^{y^2}))\\
&=U_i(y^1)-U_i(y^2).
\end{aligned}$$

□

By implementing the tolling scheme set forth in Proposition 4.1, we guarantee that all action profiles that minimize the total congestion experienced on the network are equilibria of the congestion game with tolls. However, there may be additional equilibria at which an inefficient operating condition can still occur. The following proposition establishes the uniqueness of a strict Nash equilibrium for congestion games on parallel network topologies such as the one considered in this example.

Proposition 4.2: Consider a congestion game with nondecreasing congestion functions where each driver is allowed to select any one road, i.e. $Y_i = \mathcal{R}$ for all drivers. If the congestion game has at least one strict equilibrium, then all equilibria have the same aggregate vehicle distribution over the network. Furthermore, all equilibria are strict.

Proof: Suppose action profiles y^1 and y^2 are equilibria with y^1 being a strict equilibrium. We will again use the shorthand notation $\sigma_r^{y^1}$ to represent $\sigma_r(y^1)$. Let $\sigma(y^1) := (\sigma_{r_1}^{y^1}, \ldots, \sigma_{r_n}^{y^1})$ and $\sigma(y^2) := (\sigma_{r_1}^{y^2}, \ldots, \sigma_{r_n}^{y^2})$ be the aggregate vehicle distribution over the network for equilibrium y^1 and y^2. If $\sigma(y^1) \neq \sigma(y^2)$, there exists a road a such that $\sigma_a^{y^1} > \sigma_a^{y^2}$ and a road b such that $\sigma_b^{y^1} < \sigma_b^{y^2}$. Therefore, we know that

$$\begin{aligned}
c_a\left(\sigma_a^{y^1}\right) &\geq c_a\left(\sigma_a^{y^2}+1\right)\\
c_b\left(\sigma_b^{y^2}\right) &\geq c_b\left(\sigma_b^{y^1}+1\right).
\end{aligned}$$

Since y^1 and y^2 are equilibrium with y^1 being strict

$$\begin{aligned}
c_a\left(\sigma_a^{y^1}\right) &< c_{r_i}\left(\sigma_{r_i}^{y^1}+1\right), \quad \forall r_i \in \mathcal{R}\\
c_b\left(\sigma_b^{y^2}\right) &\leq c_{r_i}\left(\sigma_{r_i}^{y^2}+1\right), \quad \forall r_i \in \mathcal{R}.
\end{aligned}$$

Using the above inequalities, we can show that

$$c_a\left(\sigma_a^{y^1}\right)\geq c_a\left(\sigma_a^{y^2}+1\right)\geq c_b\left(\sigma_b^{y^2}\right)\geq c_b\left(\sigma_b^{y^1}+1\right)>c_a\left(\sigma_a^{y^1}\right)$$

which gives us a contradiction. Therefore $\sigma(y^1) = \sigma(y^2)$. Since y^1 is a strict equilibrium, then y^2 must be a strict equilibrium as well. □

When the tolling scheme set forth in Proposition 4.1 is applied to the congestion game example considered previously, the resulting congestion game with tolls is a potential game in which no player is indifferent between distinct strategies. Proposition 4.1 guarantees us that the action profiles that minimize the total congestion experienced by all drivers on the network are

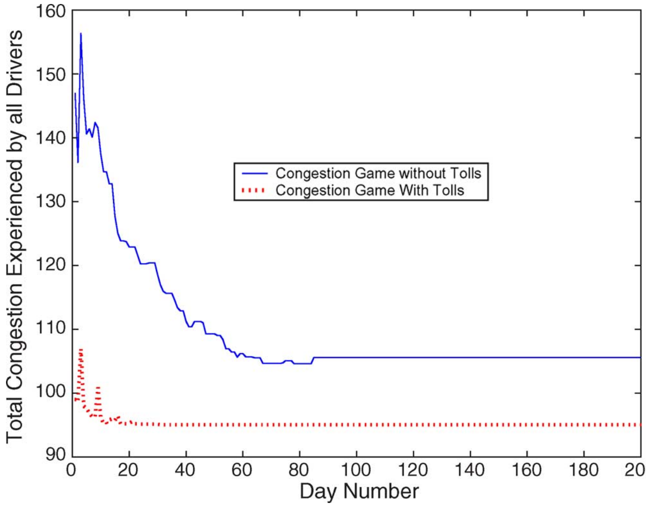

Fig. 5. Evolution of total congestion experienced by all drivers.

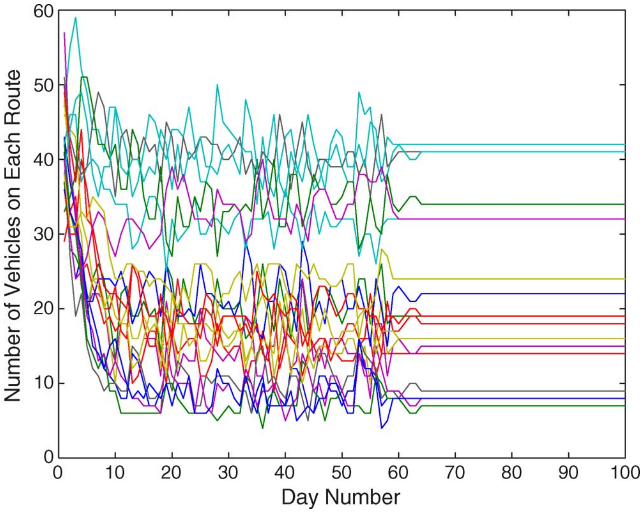

Fig. 6. Evolution of number of vehicles on each route.

in fact strict equilibria of the congestion game with tolls. Furthermore, if the new congestion functions are nondecreasing,[9] then by Proposition 4.2, all strict equilibria must have the same aggregate vehicle distribution over the network, and therefore must minimize the total congestion experienced by all drivers on the network. Therefore, the action profiles generated by fading memory JSFP with inertia converge to an equilibrium that minimizes the total congestion experienced by all users, as shown in Fig. 5.

D. Distributed Traffic Routing—General Network Topology

In this section, we will simulate the learning algorithm fading memory JSFP with inertia over a more general network topology.

Consider a congestion game with a 100 players seeking to traverse through a common network encompassing 20 different resources or roads denoted by $\mathcal{R}$. Now, an action for each player consists of multiple resources, i.e., $Y_i \subset 2^{\mathcal{R}}$. Each player's action set consists of 20 different actions chosen randomly over the set $2^{\mathcal{R}}$. Each road has a quadratic cost function with positive (randomly chosen) coefficients

$$c_{r_i}(k) = a_i k^2 + b_i k + c_i, \quad i = 1, \ldots, 20$$

where k represent the number of vehicles on that particular road. This setup can be used to model a variety of network topologies. In such a setting, a player's action sets would consist of all routes, or set of resources/roads, connecting his origin and destination.

We simulated a case where drivers choose their initial routes randomly, and every day thereafter, adjusted their routes using fading memory JSFP with inertia. The parameters $\alpha_i(t)$ are chosen as 0.5 for all days and all players, and the fading memory parameter ρ is chosen as 0.5. The number of vehicles on each road fluctuates initially and then stabilizes as illustrated in Fig. 6. The final routing profile is a Nash equilibrium.

V. Conclusion

We have analyzed 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. 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 all generalized ordinal potential games, which includes but is not limited to all congestion games, when players use some inertia either with or without exponential discounting of the historical data. The methods were 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. An interesting continuation of this research would be the case where players observe only the actual utilities they receive.

The method of proof of Theorems 2.1 and 3.1 relies on inertia to derive a positive probability of a single player seeking to make a utility improvement, thereby increasing the potential function. This suggests a convergence rate that is exponential in the game size, i.e., number of players and actions. It should be noted that inertia is simply a proof device that assures convergence for generic potential games. The proof provides just one out of multiple paths to convergence. The simulations reflect that convergence can be much faster. Indeed, simulations suggest that convergence is possible even in the absence of inertia but not necessarily for all potential games. Furthermore, recent work [35] suggests that convergence rates of a broad class of distributed learning processes can be exponential in the game size as well, and so this seems to be a limitation in the framework of distributed learning rather than any specific learning process (as opposed to centralized algorithms for computing an equilibrium). An important research direction involves characterizing the convergence rates for general multi-agent learning algorithms such as JSFP.

[9] Simple conditions on the original congestion functions can be established to guarantee that the new congestion functions, i.e congestion plus tolls, are nondecreasing.

APPENDIX

A. Proof of Theorem 2.1

The appendix is devoted to the proof of Theorem 2.1. It will be helpful to note the following simple observations:

1) The expression for $U_i(\bar{y}_i, z_{-i}(t))$ in (5) is linear in $z_{-i}(t)$.
2) If an action profile, $y^0 \in Y$, is repeated over the interval $[t, t+N-1]$, i.e.,

$$y(t) = y(t+1) = \ldots = y(t+N-1) = y^0$$

then $z(t+N)$ can be written as

$$z(t+N) = \frac{t}{t+N} z(t) + \frac{N}{t+N} \mathbf{v}^{y^0}$$

and likewise $z_{-i}(t+N)$ can be written as

$$z_{-i}(t+N) = \frac{t}{t+N} z_{-i}(t) + \frac{N}{t+N} \mathbf{v}^{y^0_{-i}}.$$

We begin by defining the quantities $\delta_i(t)$, M_u, m_u, and γ as follows. Assume that player $\mathcal{P}_i$ played a best response at least one time in the period $[0, t]$, where $t \in [0, \infty)$. Define

$$\delta_i(t) := \min\left\{0 \le \tau \le t : y_i(t-\tau) \in BR_i\left(z_{-i}(t-\tau)\right)\right\}.$$

In other words, $t - \delta_i(t)$ is the last time in the period $[0, t]$ at which player $\mathcal{P}_i$ played a best response. If player $\mathcal{P}_i$ never played a best response in the period $[0, t]$, then we adopt the convention $\delta_i(t) = \infty$. Note that

$$y_i(t-\tau) = y_i(t), \forall \tau \in \{0, 1, \ldots, \min\{\delta_i(t), t\}\}.$$

Now define

$$\begin{aligned} M_u &:= \max\left\{\left|U_i(y^1) - U_i(y^2)\right| : y^1, y^2 \in Y, \mathcal{P}_i \in \mathcal{P}\right\} \\ m_u &:= \min\left\{\left|U_i(y^1) - U_i(y^2)\right| : \left|U_i(y^1) - U_i(y^2)\right| > 0 \right. \\ &\qquad\qquad \left. y^1, y^2 \in Y, \mathcal{P}_i \in \mathcal{P}\right\} \\ \gamma &:= \lceil M_u / m_u \rceil \end{aligned}$$

where $\lceil \cdot \rceil$ denotes integer ceiling.

The proof of fading memory JSFP with inertia relied on a notion of memory dominance. This means that if the current action profile is repeated a sufficient number of times (finite and independent of time) then a best response to the weighted empirical frequencies is equivalent to a best response to the current action profile and hence will increase the potential provided that there is only a unique deviator. This will always happen with at least a fixed (time independent) probability because of the players' inertia.

In the non-discounted case the memory dominance approach will not work for the reason that the probability of dominating the memory because of the players' inertia diminishes with time. However, the following claims show that one does not need to dominate the entire memory, but rather just the portion of time for which the player was playing a suboptimal action. By dominating this portion of the memory, one can guarantee that a unilateral best response to the empirical frequencies will increase the potential. This is the fundamental idea in the proof of Theorem 2.1.

Claim 6.1: Consider a player $\mathcal{P}_i$ with $\delta_i(t) < \infty$. Let t_1 be any finite integer satisfying

$$t_1 \ge \gamma \delta_i(t).$$

If an action profile, $y^0 \in Y$, is repeated over the interval $[t, t+t_1]$, i.e.,

$$y(t) = y(t+1) = \ldots = y(t+t_1) = y^0$$

then

$$\hat{y}_i \in BR_i\left(z_{-i}(t+t_1+1)\right) \Rightarrow U_i\left(\hat{y}_i, y^0_{-i}\right) \ge U_i\left(y^0_i, y^0_{-i}\right)$$

i.e., player $\mathcal{P}_i$'s best response at time $t+t_1+1$ cannot be a worse response to y^0_{-i} than y^0_i.

Proof: Since $\hat{y}_i \in BR_i(z_{-i}(t+t_1+1))$,

$$U_i\left(\hat{y}_i, z_{-i}(t+t_1+1)\right) - U_i\left(y^0_i, z_{-i}(t+t_1+1)\right) \ge 0.$$

Expressing $z_{-i}(t+t_1+1)$ as a summation over the intervals $[0, t-\delta_i(t)-1]$, $[t-\delta_i(t), t-1]$, and $[t, t+t_1]$ and using the definition (5) leads to

$$\begin{aligned} &(t-\delta_i(t))\left[U_i\left(\hat{y}_i, z_{-i}\left(t-\delta_i(t)\right)\right) - U_i\left(y^0_i, z_{-i}\left(t-\delta_i(t)\right)\right)\right] \\ &\quad + \sum_{\tau=t-\delta_i(t)}^{t-1} \left[U_i\left(\hat{y}_i, y_{-i}(\tau)\right) - U_i\left(y^0_i, y_{-i}(\tau)\right)\right] \\ &\quad + (t_1+1)\left[U_i\left(\hat{y}_i, y^0_{-i}\right) - U_i\left(y^0_i, y^0_{-i}\right) \ge 0.\right. \end{aligned}$$

Now

$$y_i\left(t-\delta_i(t)\right) = y_i\left(t-\delta_i(t)+1\right) = \ldots = y_i(t) = y^0_i$$

and $y^0_i \in BR_i(z_{-i}(t-\delta_i(t)))$, meaning that the first term above is negative, and so

$$\begin{aligned} &\sum_{\tau=t-\delta_i(t)}^{t-1} \left[U_i\left(\hat{y}_i, y_{-i}(\tau)\right) - U_i\left(y^0_i, y_{-i}(\tau)\right)\right] \\ &\qquad + (t_1+1)\left[U_i\left(\hat{y}_i, y^0_{-i}\right) - U_i\left(y^0_i, y^0_{-i}\right) \ge 0.\right. \end{aligned}$$

This implies that

$$\left[U_i\left(\hat{y}_i, y^0_{-i}\right) - U_i\left(y^0_i, y^0_{-i}\right)\right] \ge -\frac{M_u \delta_i(t)}{t_1+1} > -m_u$$

or, alternatively

$$\left[U_i\left(y^0_i, y^0_{-i}\right) - U_i\left(\hat{y}_i, y^0_{-i}\right)\right] < m_u.$$

If the quantity in brackets were positive, this would violate the definition of m_u—unless $\hat{y}_i = y^0_i$. In either case

$$U_i\left(\hat{y}_i, y^0_{-i}\right) - U_i\left(y^0_i, y^0_{-i}\right) \ge 0.$$

□

There are certain action profile/empirical frequency values where the next play is "forced". Define the time-dependent (forced-move) set $\mathcal{F}(t) \subset Y \times \Delta(Y)$ as

$$(\bar{y}, \bar{z}) \in \mathcal{F}(t)$$
$$\Leftrightarrow$$
$$\bar{y}_i \in BR_i\left(\frac{t}{t+1}\bar{z}_{-i} + \frac{1}{t+1}\mathbf{v}^{\bar{y}_{-i}}\right), \quad \forall i \in \{1, \ldots, n\}.$$

So the condition $(y(t), z(t)) \in \mathcal{F}(t)$, implies that for all i, "today's" action necessarily lies in "tomorrow's" best response, i.e.,

$$y_i(t) \in BR_i\left(z_{-i}(t+1)\right).$$

By the rule JSFP-1, the next play $y_i(t+1) = y_i(t)$ is *forced* for all $i \in \{1, \ldots, N\}$.

Now define

$$\pi\left(t; y(t), z(t)\right) := \min\{\tau \geq 0 : (y(t+\tau), z(t+\tau)) \notin \mathcal{F}(t+\tau)\}. \tag{9}$$

If this is never satisfied, then set $\pi(t; y(t), z(t)) = \infty$.

For the sake of notational simplicity, we will drop the explicit dependence on $y(t)$ and $z(t)$ and simply write $\pi(t)$ instead of $\pi(t; y(t), z(t))$.

A consequence of the definition of $\pi(t)$ is that for a given $y(t)$ and $z(t)$, 1) $y(t)$ *must* be repeated over the interval $[t, t+\pi(t)]$. Furthermore, at time $t + \pi(t) + 1$, *at least one* player can improve (over yet another repeated play of $y(t)$) by playing a best response at time $t + \pi(t) + 1$. Furthermore, the probability that *exactly one* player will switch to a best response action at time $t + \pi(t) + 1$ is at least $\underline{\varepsilon}(1 - \bar{\varepsilon})^{n-1}$.

The following claim shows that this improvement opportunity remains even if $y(t)$ is repeated for *longer* than $\pi(t)$ (because of inertia).

Claim 6.2: Let $y(t)$ and $z(t)$ be such that $\pi(t) < \infty$. Let t_1 be any integer satisfying $\pi(t) \leq t_1 < \infty$. If

$$y(t) = y(t+1) = \ldots = y\left(t + \pi(t)\right) = \ldots = y(t + t_1)$$

then

$$y_i(t) \notin BR_i\left(z_{-i}(t + t_1 + 1)\right), \text{ for some } i \in \{1, \ldots, n\}.$$

Proof: Let $i \in \{1, \ldots, n\}$ be such that

$$y_i(t) \notin BR_i\left(z_{-i}\left(t + \pi(t) + 1\right)\right)$$

and

$$y_i(t) \in BR_i\left(z_{-i}\left(t + \pi(t)\right)\right).$$

The existence of such an i is assured by the definition of $\pi(t)$. Pick $\hat{y}_i \in BR_i(z_{-i}(t + \pi(t) + 1))$. We have

$$\begin{aligned}
&U_i\left(\hat{y}_i, z_{-i}\left(t+\pi(t)+1\right)\right) - U_i\left(y_i(t), z_{-i}\left(t+\pi(t)+1\right)\right) \\
&\quad = \left[U_i\left(\hat{y}_i, z_{-i}\left(t+\pi(t)\right)\right) - U_i\left(y_i(t), z_{-i}\left(t+\pi(t)\right)\right)\right] \\
&\qquad \times \frac{t+\pi(t)}{t+\pi(t)+1} + \left[U_i\left(\hat{y}_i, y_{-i}(t)\right) - U_i\left(y_i(t), y_{-i}(t)\right)\right] \\
&\qquad \times \frac{1}{t+\pi(t)+1} > 0.
\end{aligned}$$

Since $y_i(t) \in BR_i(z_{-i}(t + \pi(t)))$, we must have

$$U_i\left(\hat{y}_i, y_{-i}(t)\right) - U_i\left(y_i(t), y_{-i}(t)\right) > 0.$$

This implies

$$\begin{aligned}
&U_i\left(\hat{y}_i, z_{-i}(t+t_1+1)\right) - U_i\left(y_i(t), z_{-i}(t+t_1+1)\right) \\
&\quad = \left[U_i\left(\hat{y}_i, z_{-i}\left(t+\pi(t)+1\right)\right)\right. \\
&\qquad \left. - U_i\left(y_i(t), z_{-i}\left(t+\pi(t)+1\right)\right)\right] \frac{t+\pi(t)+1}{t+t_1+1} \\
&\qquad + \left[U_i\left(\hat{y}_i, y_{-i}(t)\right) - U_i\left(y_i(t), y_{-i}(t)\right)\right] \frac{t_1 - \pi(t)}{t+t_1+1} > 0.
\end{aligned}$$

□

Claim 6.3: If, at any time, $y(t)$ is not an equilibrium, then $\pi(t) \leq \gamma t$.

Proof: Let $y^0 := y(t)$. Since y^0 is not an equilibrium

$$y_i^0 \notin BR_i\left(y_{-i}^0\right), \text{ for some } i \in \{1, \ldots, n\}.$$

Pick $\hat{y}_i \in BR_i(y_{-i}^0)$ so that $U_i(\hat{y}_i, y_{-i}^0) - U_i(y_i^0, y_{-i}^0) > m_u$. If

$$y(t) = y(t+1) = \ldots = y(t + \gamma t) = y^0$$

then

$$\begin{aligned}
&U_i\left(\hat{y}_i, z_{-i}(t+\gamma t+1)\right) - U_i\left(y_i^0, z_{-i}(t+\gamma t+1)\right) \\
&\quad = \frac{t}{t+\gamma t+1}\left(U_i\left(\hat{y}_i, z_{-i}(t)\right) - U_i\left(y_i^0, z_{-i}(t)\right)\right) + \\
&\quad \frac{\gamma t+1}{t+\gamma t+1}\left(U_i\left(\hat{y}_i, y_{-i}^0\right) - U_i\left(y_i^0, y_{-i}^0\right)\right) \\
&\quad \geq \frac{-tM_u + (\gamma t+1)m_u}{t+\gamma t+1} \\
&\quad > 0.
\end{aligned}$$

□

Claim 6.4: Consider a finite generalized ordinal potential game with a potential function $\phi(\cdot)$ with player utilities satisfying Assumption 2.2. For any time $t \geq 0$, suppose that

1) $y(t)$ is not an equilibrium; and
2) $\max_{1 \leq i \leq n} \delta_i(t) \leq \bar{\delta}$ for some $\bar{\delta} \leq t$.

Define

$$\psi(t) := 1 + \max\left\{\pi(t), \gamma\bar{\delta}\right\}.$$

Then $\psi(t) \leq 1 + \gamma t$ and

$$\mathbf{Pr}\left[\phi\left(y\left(t+\psi(t)\right)\right) > \phi\left(y(t)\right) y(t), z(t) \geq \underline{\varepsilon}(1-\bar{\varepsilon})^{n(1+\gamma\bar{\delta})-1}\right.$$

and

$$\max_{1 \leq i \leq n} \delta_i\left(t + \psi(t)\right) \leq 1 + (1+\gamma)\bar{\delta}.$$

Proof: Since $y(t)$ is not an equilibrium, Claim 6.3 implies that $\pi(t) \leq \gamma t$, which in turn implies the above upper bound on $\psi(t)$.

First consider the case where $\pi(t) \geq \gamma\bar{\delta}$, i.e., $\psi(t) = 1 + \pi(t)$. According to the definition of $\pi(t)$ in (9), $y(t)$ *must* be repeated as a best response in the period $[t, t + \pi(t)]$. Furthermore, we must have

$$\max_{1 \leq i \leq n} \delta_i\left(t + \psi(t)\right) \leq 1$$

and $y_i(t) \notin BR_i(z_{-i}(t+\psi(t)))$ for at least one player $\mathcal{P}_i$. The probability that exactly one such player $\mathcal{P}_i$ will switch to a choice different than $y_i(t)$ at time $t+\psi(t)$ is at least $\underline{\varepsilon}(1-\overline{\varepsilon})^{n-1}$. But, by Claim 6.1 and no-indifference Assumption 2.2, such an event would cause

$$U_i\left(y\left(t+\pi(t)+1\right)\right) > U_i\left(y(t)\right)$$
$$\Rightarrow$$
$$\phi\left(y\left(t+\pi(t)+1\right)\right) > \phi\left(y(t)\right).$$

Now consider the case where $\pi(t) < \gamma\bar{\delta}$, i.e., $\psi(t) = 1+\gamma\bar{\delta}$. In this case

$$\max_{1\leq i\leq n} \delta_i\left(t+\psi(t)\right) \leq 1+\gamma\bar{\delta}+\bar{\delta}.$$

Moreover, the event

$$y(t) = \ldots = y(t+\gamma\bar{\delta})$$

will occur with probability at least[10] $(1-\overline{\varepsilon})^{n\gamma\bar{\delta}}$. Conditioned on this event, Claim 6.2 provides that exactly one player $\mathcal{P}_i$ will switch to a choice different than $y_i(t)$ at time $t+\psi(t)$ with probability at least $\underline{\varepsilon}(1-\overline{\varepsilon})^{n-1}$. By Claim 6.1 and no-indifference Assumption 6, this would cause

$$U_i\left(y\left(t+\psi(t)\right)\right) > U_i\left(y(t)\right)$$
$$\Rightarrow$$
$$\phi\left(y\left(t+\psi(t)\right)\right) > \phi\left(y(t)\right).$$

□

Proof of Theorem 2.1: It suffices to show that there exists a non-zero probability, $\varepsilon^* > 0$, such that the following statement holds. For any $t \geq 0$, $y(t) \in Y$, and $z(t) \in \Delta(Y)$, there exists a finite time $t^* \geq t$ such that, for some equilibrium y^*

$$\mathbf{Pr}\left[y(\tau) = y^*, \forall \tau \geq t^* | y(t), \{z_{-i}(t)\}_{i=1}^n\right] \geq \varepsilon^*. \tag{10}$$

In other words, the probability of convergence to an equilibrium by time t^* is at least ε^*. Since ε^* *does not* depend on t, $y(t)$, or $z(t)$, this will imply that the action profile converges to an equilibrium almost surely.

We will construct a series of events that can occur with positive probability to establish the bound in (10).

Let $t_0 = t+1$. All players will play a best response at time t_0 with probability at least $\underline{\varepsilon}^n$. Therefore, we have

$$\mathbf{Pr}\left[\max_{1\leq i\leq n} \delta_i(t_0) = 0 | y(t), \{z_{-i}(t)\}_{i=1}^n\right] \geq \underline{\varepsilon}^n. \tag{11}$$

Assume that $y(t_0)$ is not an equilibrium. Otherwise, according to Proposition 2.2, $y(\tau) = y(t_0)$ for all $\tau \geq t_0$.

From Claim 6.4, define t_1 and δ_1 as

$$\begin{aligned}\delta_1 &:= 1+(1+\gamma)\delta_0\\ t_1 &:= t_0+1+\max\left\{\pi(t_0), \gamma\delta_0\right\}\\ &\leq t_0+1+\gamma t_0 = 1+(1+\gamma)t_0\end{aligned}$$

where $\delta_0 := 0$. By Claim 6.4

$$\begin{aligned}\mathbf{Pr}\left[\phi\left(y(t_1)\right) > \phi\left(y(t_0)\right) | y(t_0), \{z_{-i}(t_0)\}_{i=1}^n\right]&\\ \geq \underline{\varepsilon}(1-\overline{\varepsilon})^{n(1+\gamma\delta_0)-1}&\end{aligned}$$

and

$$\max_{1\leq i\leq n} \delta_i(t_1) \leq \delta_1.$$

Similarly, for $k > 0$ we can recursively define

$$\begin{aligned}\delta_k &:= 1+(1+\gamma)\delta_{k-1}\\ &= (1+\gamma)^k\delta_0 + \sum_{j=0}^{k-1}(1+\gamma)^j\\ &= \sum_{j=0}^{k-1}(1+\gamma)^j\end{aligned}$$

and

$$\begin{aligned}t_k &:= t_{k-1}+1+\max\left\{\pi(t_{k-1}), \gamma\delta_{k-1}\right\}\\ &\leq 1+(1+\gamma)t_{k-1}\\ &\leq (1+\gamma)^k t_0 + \sum_{j=0}^{k-1}(1+\gamma)^j\end{aligned}$$

where

$$\begin{aligned}\mathbf{Pr}\left[\phi\left(y(t_k)\right) > \phi\left(y(t_{k-1})\right) | y(t_{k-1}), \{z_{-i}(t_{k-1})\}_{i=1}^n\right]&\\ \geq \underline{\varepsilon}(1-\overline{\varepsilon})^{n(1+\gamma\delta_{k-1})-1}&\end{aligned}$$

and

$$\max_{1\leq i\leq n} \delta_i(t_k) \leq \delta_k$$

as long as $y(t_{k-1})$ is not an equilibrium.

Therefore, one can construct a sequence of profiles $y(t_0), y(t_1), \ldots, y(t_k)$ with the property that $\phi(y(t_0)) < \phi(y(t_1)) < \ldots < \phi(y(t_k))$. Since in a finite generalized ordinal potential game, $\phi(y(t_k))$ cannot increase indefinitely as k increases, we must have

$$\begin{aligned}\mathbf{Pr}\left[y(t_k) \text{ is an eq. for some } t_k \in [t,\infty) | y(t), \{z_{-i}(t)\}_{i=1}^n\right]&\\ \geq \underline{\varepsilon}^n \prod_{k=0}^{|Y|-1} \underline{\varepsilon}(1-\overline{\varepsilon})^{n(1+\gamma\delta_k)-1}&\end{aligned}$$

where $\underline{\varepsilon}^n$ comes from (11). Finally, from Claim 6.1 and Assumption 2.2, the above inequality together with

$$\begin{aligned}\mathbf{Pr}\left[y(t_k) = \ldots = y(t_k+\gamma\delta_k) | y(t_k), \{z_{-i}(t_k)\}_{i=1}^n\right]&\\ \geq (1-\overline{\varepsilon})^{n\gamma\delta_k} \geq (1-\overline{\varepsilon})^{n\gamma\delta_{|Y|}}&\end{aligned}$$

implies that for some equilibrium, y^*

$$\mathbf{Pr}\left[y(\tau) = y^*, \forall \tau \geq t^* | y(t), \{z_{-i}(t)\}_{i=1}^n\right] \geq \varepsilon^*$$

[10] In fact, a tighter bound can be derived by exploiting the forced moves for a duration of $\pi(t)$.

where

$$t^* = t_{|Y|} + \gamma\delta_{|Y|} + 1 = (1+\gamma)^{|Y|} t_0 + \sum_{j=0}^{|Y|} (1+\gamma)^j$$

$$\varepsilon^* = \left(\underline{\varepsilon}^n \prod_{k=0}^{|Y|-1} \underline{\varepsilon}(1-\bar{\varepsilon})^{n(1+\gamma\delta_k)-1} \right) \left((1-\bar{\varepsilon})^{n\gamma\delta_{|Y|}} \right).$$

Since ε^* does not depend on t this concludes the proof. □

References

[1] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," in *Proc. 44th IEEE Conf. Decision Control*, Dec. 2005, pp. 6692–6697.

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

[3] H. P. Young*, Strategic Learning and Its Limits*. Oxford, U.K.: Oxford University Press, 2005.

[4] D. Fudenberg and D. Kreps, "Learning mixed equilibria," *Games Econ. Behav.*, vol. 5, no. 3, pp. 320–367, 1993.

[5] D. Monderer and A. Sela, "Fictitious Play and No-Cycling Conditions," Tech. Rep., 1997 [Online]. Available: http://www.sfb504.uni-mannheim.de/publications/dp97-12.pdf

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

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

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

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

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

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

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

[13] S. Hart and A. Mas-Colell, "Uncoupled dynamics do not lead to Nash equilibrium," *Amer. Econ. Rev.*, vol. 93, no. 5, pp. 1830–1836, 2003.

[14] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Trans. Automat. Control*, vol. 50, no. 3, pp. 312–327, Mar. 2005.

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

[16] T. Lambert, M. Epelman, and R. Smith, "A fictitious play approach to large-scale optimization," *Operations Research*, vol. 53, no. 3, pp. 477–489, 2005.

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

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

[19] S. Hart and A. Mas-Colell, "Regret based continuous-time dynamics," *Games Econ. Behav.*, vol. 45, pp. 375–394, 2003.

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

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

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

[23] D. Leslie and E. Collins, "Individual Q-learning in normal form games," *SIAM J. Control Optim.*, vol. 44, pp. 495–514, 2005.

[24] D. Leslie and E. Collins, "Generalised weakened fictitious play," *Games Econ. Behav.*, vol. 56, pp. 285–298, 2005.

[25] S. Huck and R. Sarin, "Players with limited memory," *Contrib. Theor. Econ.* vol. 4, no. 1, 2004 [Online]. Available: http://www.bepress.com/bejte/contributions/vol4/iss1/art6

[26] S. Fischer and B. Vocking, "The evolution of selfish routing," in *Proc. 12th Eur. Symp. Algorithms (ESA'04)*, 2004, pp. 323–334.

[27] S. Fischer and B. Voecking, "Adaptive routing with stale information," in *Proc. 24th Annu. ACM Symp. Principles Distrib. Comput.*, 2005, pp. 276–283.

[28] S. Fischer, H. Raecke, and B. Voecking, "Fast convergence to Wardrop equilibria by adaptive sampling methods," in *Proc. 38th Annu. ACM Symp. Theory Comput.*, 2006, pp. 653–662.

[29] A. Blum, E. Even-Dar, and K. Ligett, "Routing without regret: On convergence to Nash equilibria of regret-minimizing algorithms in routing games," in *Proc. 25th Annuu. ACM Symp. Principles Distrib. Comput.*, 2006, pp. 45–52.

[30] D. Monderer and L. S. Shapley, "Potential games," *Games Econ. Behav.*, vol. 14, pp. 124–143, 1996.

[31] I. Milchtaich, "Social optimality and cooperation in nonatomic congestion games," *J. Econ. Theory*, vol. 114, no. 1, pp. 56–87, 2004.

[32] J. G. Wardrop, "Some theoretical aspects of road traffic research," *Proc. Inst. Civil Eng.*, vol. I, pt. II, pp. 325–378, Dec. 1952.

[33] T. Roughgarden, "The price of anarchy is independent of the network topology," *J. Comput. Syst. Sci.*, vol. 67, no. 2, pp. 341–364, 2003.

[34] W. Sandholm, "Evolutionary implementation and congestion pricing," *Rev. Econ. Studies*, vol. 69, no. 3, pp. 667–689, 2002.

[35] S. Hart and Y. Mansour, "The Communication Complexity of Uncoupled Nash Equilibrium Procedures," The Hebrew University of Jerusalem, Center for Rationality, Jerusalem, Israel, Tech. Rep. DP-419, Apr. 2006.

Jason R. Marden received the B.S. and Ph.D. degrees in mechanical engineering from the University of California in Los Angeles (UCLA), in 2001 and 2007, respectively.

Since 2007, he has been a Junior Fellow in the Social and Information Sciences Laboratory, California Institute of Technology, Pasadena. His research interest is game theoretic methods for feedback control of distributed multi-agent systems.

Gürdal Arslan received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, in 2001.

From 2001 to 2004, he was an Assistant Researcher in the Department of Mechanical and Aerospace Engineering, University of California, Los Angeles. In August 2004, he joined the Electrical Engineering Department, University of Hawaii, Manoa, where he is currently an Associate Professor. His current research interests lie in the design of cooperative multi-agent systems using game theoretic methods.

Dr. Arslan received the National Science Foundation CAREER Award in May 2006.

Jeff S. Shamma received the B.S. degree from the Georgia Institute of Technology (Georgia Tech), Atlanta, in 1983 and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1988, both in mechanical engineering.

He has held faculty positions at the University of Minnesota, Minneapolis, the University of Texas at Austin, and the University of California, Los Angeles. He returned to Georgia Tech in 2007, where he is a Professor of Electrical and Computer Engineering and Julian T. Hightower Chair of Systems and Controls.