Congestion Pricing and Learning in Traffic Network Games

EMERSON MELO
California Institute of Technology

Abstract

A stochastic model describing the learning process and adaptive behavior of finitely many users in a congested traffic network with parallel links is used to prove convergence almost surely toward an efficient equilibrium for a related game. To prove this result, we assume that the social planner charges on every route the marginal cost pricing without knowing what is the efficient equilibrium. The result is a dynamic version of Pigou's solution, where the implementation is made in a decentralized way and the information about players gathered by the social planner is minimal. Our result and setting may be extended to the general case of negative externalities.

1. Introduction

Congestion in traffic networks is the classical problem of negative externalities, which is generated due to selfish routing of players. To solve this problem—in order to improve the general performance of the network—an economic solution was given in Pigou (1920), which is known as the *Pigouvian solution* for negative externalities. Pigou's solution consists in charging a toll on every route so that each player pays exactly the externality generated by its presence (in the specific route) to the other players. Once that this solution is implemented the players must consider in their choices two sources of cost: the cost due to delay in the route and the cost due to toll.

Emerson Melo, Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125 (emelos@hss.caltech.edu).

This paper was finished when the author was a researcher at the Central Bank of Chile. I warmly thank Roberto Cominetti by encourage me to write this paper. I also thank two anonymous referees and the Associate Editor for constructive suggestions that helped to improve the final presentation of the paper.

Received February 20, 2009; Accepted January 6, 2010.

Journal of Public Economic Theory, 13 (3), 2011, pp. 351–367.

This idea was called as *marginal cost pricing* and an early formalization of this one was given in Beckmann, McGuire, and Winstein (1956), where the main result says that if a social planner implements a tolls system such that the players pay exactly for the externalities that they create and if the cost functions are convex and increasing, then an "*optimal solution*"—from a social viewpoint—is reached. Despite being an effective mechanism to induce an efficient equilibrium, the efficacy of Pigou's principle rests on the assumption that the social planner knows in a precisely way that is the efficient equilibrium and according to this, he can calculate the tolls necessaries to attain that state. However, this assumption implies that the social planner has information about players, mainly respect to their preferences, without specifying how this is acquired. Behind these observations is the fact that Pigou's principle is thought as a equilibrium concept, omitting how this one may be attained. Two papers that consider dynamical processes to explain how and when the equilibrium may be reached, when the Pigouvian solution is used, are Sandholm (2005) and Sandholm (2007). Sandholm's results are built using evolutionary game theory at an aggregate population level in Sandholm (2005) and for individual players in Sandholm (2007). The main result of both papers is that if the social planner uses marginal cost pricing to correct the externalities, then the game converges toward an efficient Nash equilibrium. These results are based on the concept of potential games, which was first introduced in Rosenthal (1973) and later generalized in Monderer and Shapley (1996) and Sandholm (2001). Even though these papers prove convergence toward an efficient equilibrium, the dynamics used are postulated based on arguments of myopic behavior, which permit to get a general process of adjustment used by players. On the other hand, a simple procedure of learning used by individual players is considered in Marden et al. (2009), which is applied to congestion games in order to analyze the convergence toward the set of Nash equilibria when players' payoffs are modified by marginal cost pricing. The main characteristic of this procedure is that players only observe their own payoff obtained from the alternative chosen, that is, from the alternatives they experiment. This way of adjustment is inspired in ideas from Foster and Young (2006), and it is worth noting that this learning rule needs to assume that there exist an exogenous rate of experimentation that permits that all players can know the performance of all alternatives. Besides, the procedure works for a specific class of games called *weakly better reply*, where potential games belongs to this one. The main result given in Marden et al. (2009) is that in the long run, it is possible to stay in a Nash (efficient) equilibrium with a probability near to 1. Although the result just described is appealing, the learning procedure and its assumptions may still be considered as too restrictive in the context of traffic games with finitely many users.

In this paper, an alternative learning procedure is considered. In particular, we use the payoff-based learning procedure proposed in Cominetti, Melo, and Sorin (2010) to prove convergence toward an efficient equilibrium in the context of traffic games. This learning rule works in the following way: each player has a prior perception or estimate of the average payoff of alternative routes at network and makes an *"optimal"* decision based on this rough information by using a Logit choice rule. On this setting, we consider a social planner who modifies players' payoffs through marginal cost pricing, which is the *economic mechanism* used by this planner to reach an efficient equilibrium. Considering this fact, the payoff of the chosen alternative is then observed and is used to update the perception for that particular route. This procedure is repeated day after day, generating a discrete time stochastic process that we call the *learning process*. The basic ingredients are therefore: a state parameter; a decision rule from states to (mixed) actions; and an updating rule on the state space. Although players observe only the payoff of the specific route chosen on any given day, the observed values depend on the congestion levels determined by everybody else's choices revealing implicit information on the system as a whole. Furthermore, in the setting just described, the planner only needs to know the form of marginal cost pricing and the natural question is whether under such a simple learning mechanism based on a minimal piece of information may be sufficient to induce coordination and make the system stabilize to an efficient equilibrium.

It is necessary to make two important remarks about our paper. The first remark is about the type of learning rule that we will use. We note that our learning rule is similar to the *reinforcement model*. In fact, we use a learning procedure that preserves the qualitative features of *probabilistic choice* and *sluggish adaptation* (see Section 2 in Young 2004 for details). Despite this initial similarity, our learning process induces a specific dynamics on perceptions and strategies that appear to be structurally different from the previously studied ones (see the discussion in Cominetti, Melo, and Sorin 2010 and references therein). The second remark is about the contribution of our paper, which is mainly methodological, showing that a unified treatment is possible in which a learning process, marginal cost pricing, and a corresponding notion of equilibrium can be considered in a unified and self-consistent way, where the efficient solution may be implemented in a decentralized way by the social planner. A remarkable property of this result is that the social planner needs to gather a minimal piece of information about the game, which is appealing in the context of the traffic games with finitely many players. In addition, in this paper, we prove our convergence result using a different Lyapunov function that turns out to be different to the functional used by Cominetti, Melo, and Sorin (2010).

The paper is organized as follows. Section 2 describes the learning process in the setting of traffic games considering a social planner that modifies the payoffs through marginal cost pricing. Section 3 is dedicated to prove global convergence of the learning rule, where this result is based on the fact that the social planner's objective function is a potential for players' payoffs vector. Furthermore, we establish the link between the rest point of the learning process and the Nash equilibrium for a related game. In addition, in Section 3, we study the symmetric players case and results of local convergence are given. Finally, in Section 4, final comments and additional extensions are considered.

2. Congestion Pricing and Payoff-Based Adaptive Dynamics

The setting for the *traffic game* is as follows. Each day a set of N players, $i \in \mathcal{P}$, choose one among M alternative routes from a set $\mathcal{R}$. The combined choices of all players determine the total route loads and the corresponding travel times. Each player experiences only the cost of the route chosen on that day and uses this information to adjust the perception for that particular route, affecting the mixed strategy to be played in the next stage.

More precisely, a route $r \in \mathcal{R}$ is characterized by an increasing and convex function c_u^r that represents the average travel time of the route when it carries a load of u users, and, according to this, the congestion is captured by the following inequality $c_1^r \leq \cdots \leq c_N^r$. The set of pure strategies for each player $i \in \mathcal{P}$ is $S^i = \mathcal{R}$ and we denote Δ^i as the set of mixed strategies over S^i, where we set $\Delta = \prod_{i \in \mathcal{P}} \Delta^i$. Let $G^i(\cdot)$ the payoff function for player i and if $r_n^j \in \mathcal{R}$ denotes the route chosen by each player j at stage n, then player i's payoff is given as the negative of the experienced travel time $g_n^i = G^i(r_n) = -c_u^r$ with $r = r_n^i$ and $u = \#\{j \in \mathcal{P} : r_n^j = r\}$.

We assume that the route r_n^i is randomly chosen by player i according to a mixed strategy $\pi_n^i = \sigma^i(x_n^i) \in \Delta^i$, which depends on a vector $x_n^i = (x_n^{ir})_{r \in \mathcal{R}}$ that represents her perceptions about the payoffs of the routes available. In particular, we use the Logit model with

$$\sigma^{ir}(x^i) = \frac{\exp(\beta_i x^{ir})}{\sum_{a \in \mathcal{R}} \exp(\beta_i x^{ia})}, \tag{1}$$

where the parameter β_i has a smoothing effect when $\beta_i \downarrow 0$ leading to an uniform choice, while $\beta_i \uparrow \infty$ the probability concentrates on the pure strategy with the lower perception. As it is well known, in the traffic game just described, players' choices are made in a selfish way, that is, they do not incorporate the effect of their decisions over payoffs of other players, what implies that Nash equilibria are inefficient. This selfish routing is in opposition with the social planner's objective, who desires to maximize the aggregate welfare considering the choices of all players. In traffic games, this is equivalent to minimize the total cost of the network. Formally, let $\bar{H} : [0,1]^{\mathcal{P} \times \mathcal{R}} \to \mathbb{R}$ be defined as

$$\bar{H}(\pi) = -\,\mathbb{E}\left[\sum_{r \in \mathcal{R}} U^r c_{U^r}^r\right], \tag{2}$$

with $U^r = \sum_{i \in \mathcal{P}} X^{ir}$, where X^{ir} are independent nonhomogeneous Bernoulli's random variables with parameters $\mathbb{P}(X^{ir} = 1) = \pi^{ir}$.

In order to maximize (2), the planner charges a toll on every route, which is given by

$$p^r_u = (u-1)\left(c^r_u - c^r_{u-1}\right) \; \forall r \in \mathcal{R}, \tag{3}$$

where p^r_u is the toll charged at the route r when it carries a load of u users.

The expression (3) is a dynamic version of the classical solution of *marginal cost pricing* to correct the negative externalities generated by selfish routing of independent and noncooperative players.

This tolls system modifies the payoffs for players, so at the stage n these are given by $\tilde{g}^i_n = \tilde{G}^i(r_n) = G^i(r_n) - p^r_u = -\tilde{c}^r_u$ for every $r \in \mathcal{R}$, where $\tilde{c}^r_u = c^r_u + p^r_u$. As c^r_u is increasing and convex, it is easy to see that the inequality $\tilde{c}^r_1 \leq \tilde{c}^r_2 \cdots\cdots \leq \tilde{c}^r_N$ holds.

It is important to note that under (3) players pay for the externalities they *currently* create. The reason of this fact is due to that the social planner does not know what is the efficient equilibrium and consequently he is not able to know what tolls should be charged to attain this result. Moreover, we remark that the tolls system given by (3) is *anonymous* in the sense this one does not depend on players' identity.

Considering the previous setting and following to Cominetti, Melo, and Sorin (2010), we introduce the learning process as follows. At the stage n, the perceptions x^{ir}_n determine the choice probabilities $\pi^{ir}_n = \sigma^{ir}(x^i_n)$ that are used by each player i to select a random route $r^i_n \in \mathcal{R}$. These choices determine the load u^r_n of route r as the total (random) number of users i such that $r^i_n = r$. The payoff of route r is then given by $\tilde{g}^i_n = -\tilde{c}^r_u$ with $u = u^r_n$. At the end of the stage n, each player i observes only the cost of the chosen alternative r^i_n and updates his/her perceptions by averaging

$$x^{ir}_{n+1} = \begin{cases} (1-\gamma_n)x^{ir}_n + \gamma_n \tilde{g}^i_n & \text{if } \; r^i_n = r \\ x^{ir}_n & \text{otherwise,} \end{cases}$$

where $\gamma_n \in (0, 1)$ is a sequence of averaging factors with $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$ (a typical choice is $\gamma_n = \frac{1}{n}$). Schematically, the procedure just described looks like

$$x^{ir}_n \rightsquigarrow \pi^{ir}_n \rightsquigarrow r^i_n \rightsquigarrow u^r_n \rightsquigarrow \tilde{g}^i_n \rightsquigarrow x^{ir}_{n+1},$$

which yields a discrete time stochastic process that represents the evolution of player perceptions where the perceptions of pure strategies not played at that stage remain unchanged. We call this the *learning process* and we rewrite it in condensed form as

$$x_{n+1} - x_n = \gamma_n[w_n - x_n] \tag{4}$$

$$w^{ir}_n = \begin{cases} \tilde{g}^i_n & \text{if } \quad r^i_n = r \\ x^{ir}_n & \text{otherwise.} \end{cases} \tag{5}$$

An interesting characteristic of this learning process is that the information gathered at every stage by each player is very limited—because only the payoff of the movement realized is known—so the main question we address is whether an iterative procedure based on such a minimal piece of information can lead to coordination among the players on a steady state and how this one can be related with the social planner problem's given by (2).

Process (4) has the form of a stochastic algorithm (see Benaim 1999, Benaim, Hofbauer, and Sorin 2005) with the distribution of the random vector w_n being determined by the individual Logit rules that depend upon the prior perceptions x_n. We remark that since the route costs are bounded, the same holds for the sequences generated by (4). Hence, the asymptotic behavior of (4) can be studied by analyzing the continuous dynamics[1] of the expected movement, that is to say

$$\frac{dx}{dt} = \mathbb{E}(w \mid x) - x. \tag{6}$$

In order to make this equation more explicit we remark that if $\pi^{ir} = \mathbb{P}(X^{ir} = 1)$, then we can define the quantity

$$\overline{F}^{ir}(\pi) = \mathbb{E}\big[-\tilde{c}^r_{U^r} \,\big|\, X^{ir} = 1\big] = \mathbb{E}\big[-\big(c^r_{U^r_i+1} + p^r_{U^r_i+1}\big)\big], \tag{7}$$

with $U^r_i = \sum_{k\neq i} X^{kr}$. The expression (7) represents the average cost observed by user i when he chooses route r and the other users choose it with probabilities π^{jr} for $j \in \mathcal{P}$, $j \neq i$. Notice that $\bar{F}^{ir}(\pi)$ is a function of the probabilities $(\pi^{jr})_{j\neq i}$ only and does not depend on the probabilities with which the users choose the other arcs. Furthermore, we introduce the space of perceptions $\Omega = \prod_{i\in\mathcal{P}} \mathbb{R}^{\mathcal{R}}$ and the map $\bar{C} : \Omega \to \Omega$ that express the vector payoff as a function of the state given by

$$\bar{C}^{ir}(x) = \bar{F}^{ir}(\Sigma(x)), \tag{8}$$

where the Logit model is incorporated through the map Σ: $\Omega \to \Delta$ defined as

$$\Sigma(x) = (\sigma^i(x^i))_{i\in\mathcal{P}}, \tag{9}$$

where the latter represents the profile of mixed strategies at the state x and $\overline{F} : \Delta \to \Omega$ is the vector payoff function defined on the strategy space by $\overline{F}(\pi) = (\overline{F}^i(\pi))_{i\in\mathcal{P}}$.

LEMMA 1: *Setting* $U^r_{ij} = \sum_{k\neq i,j} X^{kr}$, *we have*

$$\overline{F}^{ir}(\pi) = \mathbb{E}\big(-c^r_{U^r_i+1}\big) + \sum_{j\neq i} \sigma^{jr}(x^j)\mathbb{E}\big(-\Delta c^r_{U^r_{ij}+2}\big).$$

[1] See Appendix for details.

Proof: As $\mathbb{E}(-(c^r_{U^r_i+1} + p^r_{U^r_i+1})) = \mathbb{E}(-c^r_{U^r_i+1}) + \mathbb{E}(-p^r_{U^r_i+1})$, we note that $\mathbb{E}(-p^r_{U^r_i+1})$ can be written as follows

$$\mathbb{E}\big(-p^r_{U^r_i+1}\big) = \mathbb{E}\big(-U^r_i\big(c^r_{U^r_i+1} - c^r_{U^r_i}\big)\big).$$
$$= \mathbb{E}\left(-\sum_{j\neq i} X^{jr} \Delta c^r_{U^r_i+1}\right).$$

As $\sum_{j\neq i} X^{jr}$ is the summation of independents nonhomogenous Bernoulli's random variables and using conditional expectation for all $j \neq i$, the preceding expression can be expressed as $\sum_{j\neq i} \sigma^{jr}(x^j)$ $\mathbb{E}(-\Delta c^r_{U^r_{ij}+2})$. ■

PROPOSITION 1: *The continuous dynamics (6) may be expressed as*

$$\frac{dx^{ir}}{dt} = \sigma^{ir}(x^i)[\overline{C}^{ir}(x) - x^{ir}], \tag{10}$$

with $\overline{C}^{ir}(x) = \mathbb{E}(-c^r_{U^r_i+1}) + \sum_{j\neq i} \sigma^{jr}(x^j)\mathbb{E}(-\Delta c^r_{U^r_{ij}+2})$.

Proof: Taking (5) into account, the expected value $\mathbb{E}(w \mid x)$ is given by

$$\mathbb{E}(w^{ir} \mid x) = \sigma^{ir}(x^i)\overline{C}^{ir}(x) + (1 - \sigma^{ir}(x^i))x^{ir}$$
$$\mathbb{E}(w^{ir} \mid x) - x^{ir} = \sigma^{ir}(x^i)[\overline{C}^{ir}(x) - x^{ir}],$$

which plugged into (6) and using Lemma 1 we get (10). ■

We shall refer to the system of differential Equation (10) as the *adaptive dynamics*. As we have noted, the learning process and adaptive dynamics have been proposed in Cominetti, Melo, and Sorin (2010), and a distinguishing feature with respect to previous work in this area is that the dynamics are not directly postulated as a mechanism of adaptive behavior, but they emerge instead as a consequence of the learning process. Besides, we observe that $\overline{C}^{ir}(x)$ does not depend on player i's perceptions x^i but it incorporates the congestion induced by all the other players.

3. Global Convergence of the Learning Process

In the traffic game setting just described, the vector payoff map $\bar{F}(\cdot)$ given by (7) may also be expressed as the gradient of a potential function. As it is well known, the traffic game is also a potential game in the sense of Monderer and Shapley (1996). However, the notion of potential that we used is closer to the one introduced in Sandholm (2001). An interesting fact is that our potential function is given by $\bar{H}(\cdot)$.

PROPOSITION 2: $\bar{F}(\pi) = \nabla \bar{H}(\pi)$ *for all* $\pi \in \Delta$.

Proof: We note that $\bar{H}(\pi) = -\sum_{r\in\mathcal{R}} \mathbb{E}[U^r c^r_{U^r}]$ and conditioning over the variables $\{X^{ir}\}_{r\in\mathcal{R}}$ we get

$$\bar{H}(\pi) = -\sum_{r\in\mathcal{R}} \left[\pi^{ir}\, \mathbb{E}\big((U^r_i + 1)c^r_{U^r_i+1}\big) + (1-\pi^{ir})\, \mathbb{E}\big(U^r_i c^r_{U^r_i}\big)\right],$$

which combined with (7) yields

$$\begin{aligned}\frac{\partial \bar{H}}{\partial \pi^{ir}}(\pi) &= -\mathbb{E}\big[(U^r_i + 1)c^r_{U^r_i+1}\big] + \mathbb{E}\big[U^r_i c^r_{U^r_i}\big] \\ &= \mathbb{E}\big[-\big(c^r_{U^r_i+1} + p^r_{U^r_i+1}\big)\big] = \bar{F}^{ir}(\pi). \qquad \blacksquare\end{aligned}$$

To study convergence of the learning process, we need some Lipschitz estimates, which can be obtained directly from Proposition 1. The following results are expressed in terms of a parameter that measures the congestion induced by an additional player, namely

$$\delta = \max\left\{c^r_u - c^r_{u-1} : r \in \mathcal{R}; u = 2, \ldots, N\right\}. \qquad (11)$$

LEMMA 2: *The second derivatives of* $\bar{H}$ *are all zero except for*

$$\frac{\partial^2 \bar{H}}{\partial \pi^{jr} \partial \pi^{ir}}(\pi) = 2\mathbb{E}\big[c^r_{U^r_{ij}+1} - c^r_{U^r_{ij}+2}\big] \in [-2\delta, 0], \qquad (12)$$

with $j \neq i$, *where* $U^r_{ij} = \sum_{k\neq i,j} X^{kr}$.

Proof: We have noted that $\frac{\partial \bar{H}}{\partial \pi^{ir}}(\pi) = \mathbb{E}[-(c^r_{U^r_i+1} + p^r_{U^r_i+1})]$ depends only on $(\pi^{kr})_{k\neq i}$. Then using Lemma 1 and conditioning on X^{jr} we get

$$\frac{\partial \bar{H}}{\partial \pi^{ir}}(\pi) = \pi^{jr}\mathbb{E}\big[-c^r_{U^r_{ij}+2}\big] + (1-\pi^{jr})\mathbb{E}\big[-c^r_{U^r_{ij}+1}\big] + \sum_{k\neq i} \pi^{kr}\mathbb{E}\big(-\Delta c^r_{U^r_{ik}+2}\big).$$

Taking partial derivative respect to π^{jr} the result follows at once. $\blacksquare$

LEMMA 3: $\|\nabla \pi^{ir}(x^i)\|_1 \leq \frac{1}{2}\beta_i$ *for all* $i \in \mathcal{P}$ *and* $x^i \in \mathbb{R}^{\mathcal{R}}$.

Proof: It suffices to note that $\frac{\partial \pi^{ir}}{\partial x^{ia}} = \beta_i \pi^{ir}(\delta_{ar} - \pi^{ia})$ with δ_{ar} equal to 1 if $a = r$ and 0 otherwise, from which we get

$$\|\nabla \pi^{ir}(x^i)\|_1 = \beta_i \pi^{ir} \sum_{a\in\mathcal{R}} |\,\delta_{ar} - \pi^{ia}\,| = 2\beta_i \pi^{ir}(1-\pi^{ir}) \leq \frac{1}{2}\beta_i. \qquad \blacksquare$$

Defining $\omega = \max_{i\in\mathcal{P}} \sum_{j\neq i} \beta_j$, we get the following Corollary.

COROLLARY 1: *For all $x, y \in \Omega$, we have*

$$\|\bar{C}(x) - \bar{C}(y)\|_\infty \leq \omega\delta\|x - y\|_\infty. \tag{13}$$

Proof: We note that for every $i \in \mathcal{P}$ and $r \in \mathcal{R}$ Equation (8) and Lemma 2 combined yield $|\bar{C}^{ir}(x) - \bar{C}^{ir}(y)| = |\frac{\partial \bar{H}}{\partial \pi^{ir}}(\Sigma(x)) - \frac{\partial \bar{H}}{\partial \pi^{ir}}(\Sigma(y))| \leq 2\delta \sum_{j\neq i}|\sigma^{jr}(x^j) - \sigma^{jr}(y^j)|$, so using Lemma 3 we conclude. ■

We are ready to state our main theorem concerning the asymptotic convergence of the learning process (4) and the adaptive dynamics (10).

THEOREM 1: *Assume in the traffic game that $\omega\delta < 1$. Then, the corresponding adaptive dynamics (10) has a unique rest point $\bar{x}$ that is a global attractor and the process (4) converges almost surely to $\bar{x}$.*

Proof: Note that if $\omega\delta < 1$ by the Corollary 1 the existence and uniqueness of $\bar{x}$ is assured, while almost sure convergence of (4) follows from global attraction and well-known results in stochastic approximation (*cf.* Benaim 1999). Hence, it suffices to show that $\bar{x}$ is an attractor by exhibiting a strict Lyapunov function with a unique minimum at $\bar{x}$. The next Lemma describes a such function. We shall use the fact that a finite maximum of smooth functions $\varphi(t) = \max\{\varphi_j(t)\colon j \in J\}$ is absolutely continuous with derivative $\dot{\varphi}(t) = \max\{\dot{\varphi}_j(t) : j \in J(t)\}$ where $J(t)$ is the set of j's at which the max is attained. ■

LEMMA 4: *If $\omega\delta < 1$ then $\Phi(x) = \max_{ir}|\dot{x}^{ir}|$ is a Lyapunov function for (10) and $\bar{x}$ is a global attractor.*

Proof: Let ir be an index where the max is attained and assume first $\dot{x}^{ir} > 0$ (recall that $\dot{x} = \frac{dx}{dt}$). Using the equality $\frac{\partial \pi^{ir}}{\partial x^{ia}} = \beta_i \pi^{ir}(\delta_{ar} - \pi^{ia})$ one gets

$$\dot{\pi}^{ir} = \beta_i \pi^{ir}\left[\dot{x}^{ir} - \sum_{a\in\mathcal{R}} \pi^{ia}\dot{x}^{ia}\right] \leq 0,$$

while (12) gives $\frac{d}{dt}[\bar{C}^{ir}(x(t))] \leq \eta\, \Phi(x)$ with $\eta = \omega\delta < 1$ so that

$$\begin{aligned}\frac{d}{dt}[\dot{x}^{ir}] &= \dot{\pi}^{ir}[\bar{C}^{ir}(x) - x^{ir}] + \pi^{ir}\frac{d}{dt}[\bar{C}^{ir}(x) - x^{ir}]\\ &\leq \pi^{ir}[\eta - 1]\ \Phi(x).\end{aligned}$$

A similar analysis holds for the case $\dot{x}^{ir} < 0$ so we deduce

$$\frac{d}{dt}\Phi(x) \leq -\min_{ir} \pi^{ir} \, [1-\eta] \, \Phi(x).$$

Now, since $\bar{C}^{ir}(x) \in [-\tilde{c}^r_N, -\tilde{c}^r_1]$ it follows easily from (10) that $x(t)$ remains bounded and therefore π^{ir} stays away from 0 so that $\frac{d}{dt}\Phi(x(t)) \leq -\epsilon \, \Phi(x(t))$ for some $\epsilon > 0$. This implies that Φ is a Lyapunov function that decreases to 0 exponentially fast along the trajectories of (10), and since $\bar{x}$ is the unique point with $\Phi(\bar{x}) = 0$ the conclusion follows. ■

It is important to make some remarks about our convergence result. First, Theorem 1 establishes global convergence of the learning process without considering the existence of a related game. As a matter of fact, to derive the dynamic (10), we do not need to assume that players know that they are involved in a game. Moreover, players do not need to gather information about actions or payoff functions of other players. However, in the next section, we shall establish the link between the rest points of (10) and Nash equilibrium for a related game. A second remark is about the condition of convergence, where we note that in traffic games a standard assumption is that an individual player has a negligible effect on the payoff of a specific route, what in our model is captured by a small parameter δ. Thus, our condition $\omega\delta < 1$ can be considered as suitable in the context of traffic games with finitely many players. Likewise, the condition $\omega\delta < 1$ can be strengthen to the class of random utility models satisfying the gradient condition $\|\nabla\pi^{ir}(x^i)\|_1 \leq K_i$, with $K_i > 0$ for all $i \in \mathcal{P}$ and $x^i \in \mathbb{R}^{\mathcal{R}}$, where for the Logit choice rule this condition is given by Lemma 4. This point is important because we are not constrained to consider homogeneous choice rules for players. Finally, we must mention that an alternative proof of Theorem 1 can be given using our Corollary 2 combined with Theorem 2 in Cominetti, Melo, and Sorin (2010).[2]

3.1. Rest Points and Nash Equilibrium

From general results on stochastic algorithms, we know that the rest points of the continuous dynamics (10) are natural candidates to be limit points for the stochastic process (4). Since $\sigma^{ir}(x^i) > 0 \, \forall r \in \mathcal{R}$, these rest points are the fixed points of the map $x \mapsto \bar{C}(x)$ whose existence follows easily from Brouwer's Theorem if one notes that this map is continuous with bounded range. We denote $\mathcal{E}$ as the set of rest points for (10). As was noted in Cominetti, Melo, and Sorin (2010), there is a correspondence one-to-one between rest points and the associated π's that can be associated with the concept of *quantal response equilibria* introduced in McKelvey and Palfrey (1995). Formally, we get

[2] In particular, Theorem 2 in Cominetti, Melo, and Sorin (2010) provides a convergence result for a general class of games.

PROPOSITION 3: *The map* $x \longmapsto \Sigma(x)$ *is one-to-one over the set* $\mathcal{E}$.

Proof: The proof is straightforward if we note that the fixed point equation $x = \bar{C}(x)$ can be restated as a coupled system in (x, π)

$$\begin{cases} \pi = \Sigma(x) \\ x = \bar{F}(\pi). \end{cases}$$

Then for $x \in \mathcal{E}$ the map, $x \longmapsto \Sigma(x)$ has an inverse image given by $\pi \longmapsto \bar{F}(\pi)$. ■

We use Proposition 2 to establish a link between the set $\mathcal{E}$ and the Nash equilibrium for a related N-person game $\mathcal{G}$ defined by strategy sets $S^i = \Delta(\mathcal{R})$ for all $i \in \mathcal{P}$ and payoff function $\mathcal{G} : \bigotimes_{i\in\mathcal{P}} S^i \to \mathbb{R}^N$ given by

$$\mathcal{G}^i(\pi) = \langle \pi^i, \bar{F}^i(\pi) \rangle - \frac{1}{\beta_i} \sum_{r\in\mathcal{R}} \pi^{ir}[\ln \pi^{ir} - 1],$$

which is a congestion game perturbed by an entropy term.

THEOREM 2: *Consider* $\omega\delta < 1$. *Then,* $\pi = \Sigma(\mathcal{E})$ *is the Nash equilibrium of the perturbed game* $\mathcal{G}$. *Moreover, this equilibrium is unique.*

Proof: By Proposition 3 in Cominetti, Melo, and Sorin (2010), we know that $\Sigma(\mathcal{E})$ is the set of Nash equilibria for $\mathcal{G}$. Furthermore, if $\omega\delta < 1$, by Theorem 1 we get $\mathcal{E} = \{\bar{x}\}$. Finally, as we just noted, there exists a correspondence one-to-one between rest point $\bar{x}$ and the associated π, so the uniqueness of Nash equilibrium for $\mathcal{G}$ it follows. ■

The Theorem 3 says that at the Nash equilibrium each player considers the effect that his/her decision has upon payoffs of other players. To reach this result, the planner only needs to charge the *current* tolls that are given by (3). Furthermore, it is worth noting that the efficient result is attained in a decentralized way, where the planner does not need to take into account players' identity.

Remark 1: Payoff-based learning rules have been proposed in Marden and Shamma (2009) and Young (2009). However, their results differ with our approach in two important aspects. First, the cited papers study payoff-based learning exploiting the theory of perturbed Markov chains (in discrete time), which turns out to be structurally different to our approach based on stochastic approximation. In particular, Marden and Shamma (2009) and Young (2009) study the stochastically stable sets for the perturbed Markov process induced by their learning rules, whereas we study the rest points for the continuous and deterministic dynamics given by (10). The second difference is about the application of our learning rule in congestion games. We note the models proposed in Marden and Shamma (2009) and Young (2009) can also be applied to traffic games; however, in our results, we provide an explicit condition depending on the parameter δ. This condition is appealing because our model shows how the negligible effect assumption in the context of traffic games is fundamental to attain convergence.

3.2. The Symmetric Case

In this section, we assume $\beta_i \equiv \beta\, \forall i \in \mathcal{P}$ and according to this (1) is given by a common Logit function that we denote as $\sigma(\cdot)$. Under this assumption is reasonable to expect that a rest point is a situation where all players share the same perceptions, i.e. : $x^i = x^j$ for all $i, j \in \mathcal{P}$. In fact, when $\beta\delta$ is small, then only and only one rest point is symmetric.

LEMMA 5: *For all $x, y \in \Omega$, each $i, j \in \mathcal{P}$ and every $r \in \mathcal{R}$, we have*

$$|\bar{C}^{ir}(x) - \bar{C}^{jr}(x)| \leq \beta\delta \, \|x^i - x^j\|_\infty. \tag{14}$$

Proof: We observe that the only difference between $\bar{F}^{ir}$ and $\bar{F}^{jr}$ is an exchange of π^{ir} and π^{jr}. Besides, by Lemma 1, we know $\bar{F}^{ir}(\pi) = \mathbb{E}(-c^r_{U^r_i+1}) + \sum_{k\neq i} \sigma^r(x^k)\mathbb{E}(-\Delta c^r_{U^r_{ik}+2})$. Thus, Proposition 4 and Lemma 4 combined imply that $|\bar{F}^{ir}(\pi) - \bar{F}^{jr}(\pi)| \leq 2\delta\,|\pi^{ir} - \pi^{jr}|$ and then (14) follows from the equality $C(x) = F(\Sigma(x))$ and Lemma 5. ■

The existence and uniqueness of a symmetric rest point has been established in Cominetti, Melo, and Sorin (2010). The following proposition establishes the symmetry of a rest point for (10) when $\beta\delta < 1$.

PROPOSITION 4: *Let $\beta_i \equiv \beta$ for all $i \in \mathcal{P}$. If $\beta\delta < 1$ then every rest point of (10) is symmetric and unique.*

Proof: As we just noted the existence and uniqueness of a symmetric rest point it follows from Theorem 3 in Cominetti, Melo, and Sorin (2010). To prove symmetry consider us that $\beta\delta < 1$ and let x be any rest point. For any two players $i, j \in \mathcal{P}$ and all routes $r \in \mathcal{R}$, Lemma 5 gives

$$|x^{ir} - x^{jr}| = |\bar{C}^{ir}(x) - \bar{C}^{jr}(x)| \leq \beta\delta\|x^i - x^j\|_\infty$$

and then $\|x^i - x^j\|_\infty \leq \beta\delta\|x^i - x^j\|_\infty$ that implies $x^i = x^j$. ■

COROLLARY 2: *If $\beta_i \equiv \beta$ for all $i \in \mathcal{P}$ then the game $\mathcal{G}$ has a unique symmetric equilibrium. Moreover, if $\beta\delta < 1$ then every equilibrium is symmetric (hence unique).*

It is interesting to note that in the symmetric case the condition of stability of rest points given in Theorem 1 becomes in $\beta\delta < \frac{1}{N-1}$, which is more and more exigent as the number of players increase. However, we can obtain a local result which only depends on $\beta\delta < 1$. In fact, this result is a slight variation of the Theorem 3 in Cominetti, Melo, and Sorin (2010).

THEOREM 3: *If $\beta_i \equiv \beta$ for all $i \in \mathcal{P}$ with $\beta\delta < 1$ then the unique rest point $\hat{x} = (\hat{y}, \dots, \hat{y})$ is symmetric and a local attractor for the adaptive dynamics (10).*

Proof: See proof of Theorem 3 in Cominetti, Melo, and Sorin (2010) and considering $\beta\delta < 1$ instead of $\beta\delta < 2$. ■

4. Comments and Final Remarks

Two important remarks can be made for the specific model here considered. The first remark is based on the observation made in Cole, Dodis, and Roughgarden (2006) about the possibility of considering an alternative tolls scheme that is not based on marginal cost pricing. The argument to consider an alternative tolls scheme is because if the social planner has a objective function as (2), then the disutility for players generated by marginal cost pricing is not considered (for details see 3 in Cole, Dodis, and Roughgarden 2006). This observation implies the study different forms of pricing that could be better than marginal cost pricing. We can adapt our model to analyze this issue. In fact, let us consider a nondecreasing function p_u^r that represents the toll to be charged when route $r \in \mathcal{R}$ carries a load of u players with $p_1^r \leq \cdots \leq p_N^r$. So, this tolls system implies that players' payoffs at the stage n are given by $\tilde{g}_n^i = \tilde{G}^i(r_n) = G^i(r_n) - p_u^r = -\tilde{c}_u^r$, for all $r \in \mathcal{R}$, where it follows $\tilde{c}_1^r \leq \cdots \leq \tilde{c}_N^r$. Under this tolls scheme, the result in Proposition 1 does not hold any longer. However, we can recover the convergence results if we define the function given by: $H(\pi) = -\mathbb{E}(\sum_{r\in\mathcal{R}} \sum_{u=1}^{U^r} \tilde{c}_u^r)$. This function is a potential for $\bar{F}(\cdot)$, namely, $\nabla H(\pi) = \bar{F}(\pi) \forall \pi \in \Delta$, where the argument to prove it is the same one used in the proof of Proposition 1. Although we cannot relate this potential function with social planner's problem, this result permits to attain convergence of the learning process for the case of a general nondecreasing function p_u^r. In terms of Cole, Dodis, and Roughgarden (2006), we may study the case when the social planner has the objective function: $\tilde{H}(\pi) = -\mathbb{E}(\sum_{r\in\mathcal{R}} U^r \tilde{c}_{U^r}^r) = -\mathbb{E}(\sum_{r\in\mathcal{R}} U^r (c_{U^r}^r + p_{U^r}^r))$, that is, the case when the planner considers both the disutility due to delay and the disutility generated by tolls.

The second remark is about the case when each player has *private valuations* for every route. More precisely, let us consider the following payoff function: $G^i(r_n) = \nu^{ir} - c_u^r$ for all $i \in \mathcal{P}$, $r \in \mathcal{R}$, where the parameter ν^{ir} represents player i's valuation for using route r, so it varies among players and routes.[3] Under this setting for payoffs, the adaptive dynamics (10) is the same and (2) is still a potential function for $\bar{F}(\cdot)$. Thus, the results of convergence are invariant.

Finally, an interesting extension of our results would be to study the case when each route is owned by a monopolist that must charge prices in an optimal way (in order to maximize its profit). This issue raises the question about how players and a monopolist could learn simultaneously and how the equilibrium would be attained.

Appendix: Stochastic Approximation

In this appendix, we briefly revise the theory of stochastic approximation. Our review is based on Benaim (1999).

The following definition establishes what it is understood by stochastic algorithm.

DEFINITION 1: *Let $\{x_n\}_{n\in\mathbb{N}}$ be a discrete time process living in $\mathbb{R}^m$. We say that this process is a stochastic algorithm if it can be written as:*

$$x_{n+1} - x_n = \gamma_{n+1}(F(x_n) + U_{n+1}), \tag{A1}$$

where

- $F : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ *is a continuous map.*
- $\{\gamma_n\}_{n\geq 1}$ *is a given sequence of nonnegative numbers such that*

$$\sum_k \gamma_k = \infty \quad y \quad \lim_{n\longrightarrow\infty} \gamma_n = 0.$$

- $U_n \in \mathbb{R}^m$ *are (deterministic or random) perturbations.*

Intuitively, the idea of stochastic approximation is to compare the behavior of a sample path $\{x_n\}_{n\in\mathbb{N}}$ with the trajectories induced (in continuous time) by the vector field F.

In particular, the vector field F is said to be *globally integrable* if it has unique integral curves. A well-known example is the case when F is a locally Lipschitz vector field, which is always globally integrable. The following result is fundamental in the theory of stochastic approximation.

PROPOSITION 5 (Benaim 1999): *Let F be a continuous globally integrable vector field. Assume that:*

[3] This framework has been considered by Miltaich (2004) in the context of congestion games with a continuum of players.

A1 *For all* $T > 0$

$$\lim_{n \longrightarrow \infty} \sup_{k} \left\{ \left\| \sum_{i=n}^{k-1} \gamma_{i+1} U_{i+1} \right\| : k = n+1, \ldots\ldots, m(\tau_n + T) \right\} = 0$$

or equivalently

$$\lim_{t \longrightarrow \infty} \Delta(t, T) = 0$$

with

$$\Delta(t, T) = \sup_{0 \leq h \leq T} \left\| \int_t^{t+h} \bar{U}(s)\, ds \right\| \tag{A2}$$

A2 $\sup_n \|x_n\| < \infty$ *, or*

A2' *F is Lipschitz and bounded on a neighborhood of* $\{x_n \colon n \geq 0\}$.

Then, the interpolated process X is an asymptotic pseudo trajectory for the flow Φ *induced by F. Furthermore, under the assumption A2', for* $t \geq 0$ *large enough we have the estimate*

$$\sup_{0 \leq h \leq T} \|X(t+h) - \Phi_h(X(t))\| \leq C(T)[\Delta(t-1, T+1) + \sup_{t \leq s \leq t+T} (\bar{\gamma}(s))], \tag{A3}$$

where $C(T)$ *is a constant depending only on T and F.*

Proof: See Benaim (1999) pp. 13–14. ■

Because of this result, we can study the discrete time process (15) using the following dynamical system

$$\dot{x} = F(x). \tag{A4}$$

The powerful of this approach is given by the fact we can study all convergence properties of (15) using (18), which in some cases turn out to be easier. In applications of Proposition 5, when U_n is random, one usually tries to verify assumption *A1* holds with probability 1. Formally, in the stochastic case, we have the following definition:

Definition 2: *Let* $\{\Omega, \mathcal{F}, P\}$ *be a probability space and* $\mathcal{F}_n$ *a nondecreasing sequence of* σ-algebras of $\mathcal{F}$. *We say that a stochastic process* $\{x_n\}_{n \in \mathbb{N}}$ *given by (15), satisfies the Robbins-Monro conditions if:*

- $\{\gamma_n\}_{n \in \mathbb{N}}$ *is a deterministic sequence.*
- $\{U_n\}_{n \in \mathbb{N}}$ *is adapted:* U_n *is measurable with respect to* $\mathcal{F}_n$ *for each* $n \geq 0$
- $\mathbb{E}(U_{n+1} \mid \mathcal{F}_n) = 0$

PROPOSITION 6 (Benaim 1999): *Let* $\{x_n\}_{n\in\mathbb{N}}$ *given by (15) be a Robbins-Monro algorithm. Suppose that for some* $q \geq 2$ *such that*

$$\sup_n \mathbb{E}(\| U_{n+1} \|^q) < \infty$$

and

$$\sum_n \gamma_n^{1+q/2} < \infty.$$

Then, the assumption **A1** *of Proposition 6 holds with probability 1.*

The previous Proposition allow us to study the learning process (4) using its deterministic and continuous time version.

References

BECKMANN, M., C. McGUIRE, and C. WINSTEIN (1956) *Studies in the Economics of Transportation*. New Haven, CT: Yale University Press.

BENAIM, M. (1999) Dynamics of stochastic approximation algorithms, *in Séminaire de Probabilités, Lecture Notes in Math 1709*, Springer, Berlin 1-68.

BENAIM, M., J. HOFBAUER, and S. SORIN (2005) Stochastic approximations and differential inclusions, *SIAM Journal of Control Optimization* **44**, 328–348.

COLE, R., Y. DODIS, and T. ROUGHGARDEN (2006) How much can taxes help selfish routing, *Journal of Computer and System Sciences* **72**, 444–467.

COMINETTI, R., E. MELO, and S. SORIN (2010) A payoff-based learning procedure and its application to traffic games, *Games and Economic Behavior* **70**, 71–83.

FOSTER, D., and H. P. YOUNG (2006) Regret testing: Learning to play nash equilibrium without knowing you have an opponent, *Theoretical Economics* **1**(1), 341–367.

MARDEN, J. R., and J. S. SHAMMA (2009) Revisiting log-linear learning: Asynchrony, completeness and payoff based implementation, *Working Paper*.

MARDEN, J. R., H. P. YOUNG, G. ARSLAN, and J. SHAMMA (2009) Payoff based dynamics for multiplayer weakly acyclic games, *SIAM Journal on Control and Optimization* **48**, 373–396.

McKELVEY, R., and T. PALFREY (1995) Quantal response equilibria for normal form games, *Games and Economic Behavior* **10**, 6–38.

MILTAICH, I. (2004) Social optimality and cooperation in nonatomic congestion games, *Journal of Economic Theory* **114**, 56–87.

MONDERER, D., and L. SHAPLEY (1996) Potential games, *Games and Economic Behavior* **14**, 124–143.

PIGOU, A. (1920) *The economic of welfare*. London: Macmillan.

ROSENTHAL, R. (1973) A class of games possessing pure-strategy nash equilibria, *International Journal of Game Theory* **2**, 65–67.

SANDHOLM, W. (2001) Potential games with continuous player sets, *Journal of Economic Theory* **97**, 81–108.

SANDHOLM, W. (2005) Negativies externalities and evolutionary implementation, *Review of Economic Studies* **72**, 885–915.

SANDHOLM, W. (2007) Pigouvian pricing and stochastic evolutionary implementation, *Journal of Economic Theory* **132**, 47–69.

YOUNG, H. P. (2004) *Strategic Learning and Its Limits*, New York: Oxford University Press.

YOUNG, H. P. (2009) Learning by trial and error, *Games and Economic Behavior* **65**, 626–643.