On the convergence of no-regret learning in selfish routing

Walid Krichene WALID@EECS.BERKELEY.EDU
University of California, 652 Sutardja Dai Hall, Berkeley, CA 94720 USA

Benjamin Drighès BENJAMIN.DRIGHES@POLYTECHNIQUE.EDU
Ecole Polytechnique, Route de Saclay, 91120 Palaiseau, France

Alexandre Bayen BAYEN@BERKELEY.EDU
University of California, 642 Sutardja Dai Hall, Berkeley, CA 94720 USA

Abstract

We study the repeated, non-atomic routing game, in which selfish players make a sequence of routing decisions. We consider a model in which players use regret-minimizing algorithms as the learning mechanism, and study the resulting dynamics. We are concerned in particular with the convergence to the set of Nash equilibria of the routing game. No-regret learning algorithms are known to guarantee convergence of a subsequence of population strategies. We are concerned with convergence of the actual sequence. We show that convergence holds for a large class of online learning algorithms, inspired from the continuous-time replicator dynamics. In particular, the discounted Hedge algorithm is proved to belong to this class, which guarantees its convergence.

1. Introduction

Routing games are important in modeling and understanding the interaction of non-cooperative players who share resources, such as roads in a road network and links in a communication network. They have been studied extensively, including the seminal work of Beckmann et al. (1955) and Dafermos & Sparrow (1969). In a one-shot scenario, selfish players choose the routes that minimize their individual travel time. One solution concept to the game is the Nash equilibrium, also called Wardrop equilibrium in the traffic literature (1952). In some classes of games, Nash equilibria can be hard to compute and have been questioned as a realistic equilibrium concept, for example by Papadimitriou (1994). By contrast, for one-shot non-atomic routing games, Nash equilibria are known to be easy to compute as they can be expressed as the solution to a convex optimization problem, using a convex potential function, due to Rosenthal (1973). This is an argument in favor of the one-shot routing game model. However, most realistic scenarios do not correspond to a one-shot game, but rather a repeated game, in which players make a sequence of routing decisions and may adapt their strategies given the outcome on previous days. Therefore studying the repeated routing game is important to understand how players can *arrive at the equilibrium*. Arguably, a good learning model for the population of players should be distributed and easy to implement by individual players. A natural framework is that of online learning.

No-regret learning is of particular interest, given its generality and ease of implementation, and the fact that it only requires the current losses to be revealed. The Hedge algorithm is one example of no-regret learning, introduced for Machine Learning by Freund & Schapire (1999), a generalization of the weighted majority algorithm of Littlestone & Warmuth (1989). Cesa-Bianchi & Lugosi (2006) give convergence results, together with convergence rates, for no-regret algorithms. These results hold for a broad class of games. However, they guarantee convergence of the *time-averaged strategies*, and not the actual sequence of strategies.

Other learning processes have been studied for repeated routing games, such as fictitious play by Monderer & Shapley (1996), adaptive sampling by Fischer et al. (2010) or continuous-time replicator dynamics by Fischer & Vöcking (2004), which is also of particular interest in evolutionary game theory, see for ex-

Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

ample Weibull (1997). For some classes of continuous-time dynamics, convergence of the actual sequence is guaranteed. For example, Sandholm proved convergence in continuous-time potential games for a class of evolutionary dynamics which satisfy a positive correlation condition (2001). Blum et al. proved in (2006) that under no-regret learning, the resulting sequence of population strategies converges, on a $1-\epsilon$ fraction of days, to the set of ϵ-approximate Nash equilibria, and they gave explicit convergence rates that depend on the Lipschitz constant of the latency functions.

We are concerned with convergence of the actual sequence of strategies (as opposed to a subsequence). Our approach combines ideas from evolutionary dynamics and no-regret learning. In Sections 2 and 3, we present the model and summarize existing results. In Section 4, we prove that convergence holds for a class of algorithms, which have sublinear discounted regret, and which can be viewed as a generalization of replicator dynamics.

2. The model

2.1. Non-atomic routing game

We consider a set $\mathcal{X}$ of players. A *routing game* is a non-cooperative game played on a directed graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ representing a network, and in which a pure strategy corresponds to a directed path on the graph. To formalize the notion of non-atomicity, we endow $\mathcal{X}$ with a structure of measurable space $(\mathcal{X},\mathcal{M},m)$, where $\mathcal{M}$ is a σ-algebra of measurable sets, and m is a measure. The set of players is said to be *non-atomic* if each single player $x\in\mathcal{X}$ is negligible for m.

We consider a setting similar to (Wardrop, 1952), in which the set of players is partitioned in *populations* or *commodities* $\mathcal{X}=\sqcup_{k=1}^{K}\mathcal{X}_k$, where each $\mathcal{X}_k$ is measurable and has positive finite measure. Formally, the model is defined by the tuple $(\mathcal{E};K;(\mathcal{X}_k)_{k\in[K]};(\mathcal{P}_k)_{k\in[K]};(c_e)_{e\in\mathcal{E}})$, where $\mathcal{E}$ denotes the finite set of edges, K is the number of commodities, $[K]$ denotes the set $\{1,\dots,K\}$, and for all k, $\mathcal{P}_k\subseteq P(\mathcal{E})$ is a set of *paths* (*pure strategies*) available to players in $\mathcal{X}_k$. For each k, all paths in $\mathcal{P}_k$ have a common source $s_k\in\mathcal{V}$ and a common destination $t_k\in\mathcal{V}$. We will denote $\mathcal{P}$ the disjoint union $\mathcal{P}=\sqcup_{k=1}^{K}\mathcal{P}_k$. The positive measure of $\mathcal{X}_k$ is denoted $F_k=m(\mathcal{X}_k)$ and called the total flow of $\mathcal{X}_k$. For each edge e, $c_e(\cdot):\mathbb{R}_+\to\mathbb{R}_+$ is an edge latency function satisfying the following assumption:

Assumption 1. *The latency functions c_e are assumed to be continuous, non-decreasing, and locally Lipschitz.*

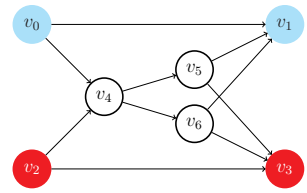

Figure 1. Example of a network with $K=2$ populations. Population 1 travels from v_0 to v_1, and population 2 travels from v_2 to v_3.

At a microscopic scale, the joint action of all players in $\mathcal{X}$ can be represented by an action profile $A\colon x\in\mathcal{X}\mapsto A(x)$ which maps a player $x\in\mathcal{X}_k$ to a path $A(x)\in\mathcal{P}_k$. This function is assumed to be $\mathcal{M}$-measurable, and defines, for each population $\mathcal{X}_k$, a *path distribution* $\mu^k=(\mu^k_p)_{p\in\mathcal{P}_k}$, where $\mu^k_p=\frac{1}{F_k}\int_{\mathcal{X}_k}\mathbb{1}_{\{A(x)=p\}}dm(x)$ is the fraction of players utilizing path $p\in\mathcal{P}_k$. We have $\mu^k\in\Delta^{\mathcal{P}_k}$, the simplex on $\mathcal{P}_k$, that is, $\Delta^{\mathcal{P}_k}=\{u\in\mathbb{R}_+^{\mathcal{P}_k}\colon\sum_{p\in\mathcal{P}_k}u_p=1\}$. The pure strategy profile can be summarized at a macroscopic scale using the product of distributions $\mu=(\mu^1,\cdots,\mu^K)\in\Delta^{\mathcal{P}_1}\times\cdots\times\Delta^{\mathcal{P}_K}$. The product of simplexes will be denoted Δ.

The distribution μ determines the edge flow or *load*, defined as $\phi_e=\sum_{k=1}^{K}F_k\sum_{p\in\mathcal{P}_k:e\in p}\mu^k_p$. This can be written compactly as $\phi_e=(M\mu)_e$ where $M=[M^1|\dots|M^K]\in\mathbb{R}^{\mathcal{E}\times\mathcal{P}}$ is a weighted incidence matrix:

$$\forall e\in\mathcal{E},\ \forall k\in[K],\ \forall p\in\mathcal{P}_k\quad M^k_{e,p}=\begin{cases}F_k & \text{if } e\in p\\ 0 & \text{otherwise}\end{cases}$$

For each edge e, the edge load determines the edge latency, given by $c_e(\phi_e)$. Finally, the loss of a player who chooses path p is simply the sum of edge latencies along the path, $\sum_{e\in p}c_e(\phi_e)$. This latency is entirely determined by the distribution μ, so we define a path latency function (or loss function) $\ell_p:\mu\in\Delta\mapsto\ell^k_p(\mu)=\sum_{e\in p}c_e((M\mu)_e)$. Finally, we write $\ell^k(\mu)$ to denote the vector of path latencies $(\ell^k_p(\mu))_{p\in\mathcal{P}_k}$, and $\ell(\mu)=(\ell^1(\mu),\dots,\ell^K(\mu))$.

2.2. Nash equilibria and the Rosenthal potential function

Given this setting, we now define Nash equilibria of the routing game.

Definition 2.1 (Nash equilibrium)**.** *A distribution $\mu\in\Delta$ is a Nash equilibrium if for every population k, whenever $\mu^k_p>0$ for some path $p\in\mathcal{P}_k$, then $\ell^k_{p'}(\mu)\geq\ell^k_p(\mu)$ for all $p'\in\mathcal{P}_k$. We will denote by $\mathcal{N}\subset\Delta$ the set of Nash equilibria.*

The definition implies that, for a commodity k, all

paths with non-zero mass have equal latencies and paths with zero mass have larger latencies.

There is a natural potential function that allows one to formulate the problem of computing the set $\mathcal{N}$ of Nash equilibria as the solution of a convex optimization problem. Consider the function

$$V(\mu) = \sum_{e \in E} \int_0^{(M\mu)_e} c_e(u)du \tag{1}$$

The gradient of V is the vector of path latencies:

$$\forall k,\ \forall p \in \mathcal{P}_k, \quad \frac{\partial V}{\partial \mu_p^k}(\mu) = F_k \ell_p^k(\mu) \tag{2}$$

Theorem 1. *(Rosenthal, 1973) $\mathcal{N}$ is the set of minimizers of V in Δ. It is a non-empty convex compact set. We denote $V_{\mathcal{N}}$ the value of V on $\mathcal{N}$.*

A proof can be found for example in (Roughgarden, 2007). As a result of Theorem 1, computing the Nash equilibria of the routing game can be done efficiently by minimizing the potential. However, the idea of minimizing a potential function cannot be directly applied to designing a distributed learning algorithm, as it would a priori require coordination between players.

2.3. Restricted Nash equilibria

In the analysis, we use a weaker notion of equilibrium, introduced by Fischer & Vöcking in (2004).

Definition 2.2 (Restricted Nash equilibrium)**.** *A product distribution μ is a restricted Nash equilibrium if all paths with non-zero mass have equal latencies for each commodity i.e. for all k and all $p, p' \in \mathcal{P}_k$ such that $\mu_p^k, \mu_{p'}^k > 0$, $\ell_p^k(\mu) = \ell_{p'}^k(\mu)$. We will denote $R\mathcal{N}$ the set of restricted Nash equilibria.*

Such an equilibrium is restricted in the sense that it would be a Nash equilibrium of the routing game if we restricted the set of paths to its support (Fischer & Vöcking, 2004).

Remark 1. *Restricted Nash equilibria are also minimizers of the potential function V if we restrict the feasible set to distributions with the same support. As the number of supports is finite, the set $V(R\mathcal{N})$ of potential values of restricted Nash equilibria is also finite.*

3. No-regret learning in the repeated routing game

3.1. The online learning framework

We now define the online learning setting. Assume players make decisions repeatedly, and index iterations by $\tau \in \mathbb{N}$. For each commodity k, every player $x \in \mathcal{X}_k$ maintains a *mixed strategy* $\pi^{(\tau)}(x) \in \Delta^{\mathcal{P}_k}$, which reflects her preferences on paths, and randomly draws a path $A^{(\tau)}(x) \sim \pi^{(\tau)}(x)$. Similarly to the one-shot case, the path profile $A^{(\tau)}$ defines a distribution $\mu^{(\tau)} \in \Delta$.

Algorithm 1 Online learning setting

Input: For every player $x \in \mathcal{X}_k$, a learning algorithm $(h_\tau^x)_\tau$ and initial distribution $\pi^{(0)}(x) \in \Delta^{\mathcal{P}_k}$.

1: **for** each time step τ **do**
2: Every player x independently draws a path $A^{(\tau)}(x) \sim \pi^{(\tau)}(x)$.
3: For all k, the vector of path losses $\ell^k(\mu^{(\tau)})$ is revealed to players in $\mathcal{X}_k$. Players incur losses corresponding to their path choice.
4: Every player updates her strategy: $\pi^{(\tau+1)}(x) = h_\tau^x\left((\ell^k(\mu(t)))_{t\le\tau}, \pi^{(\tau)}(x)\right)$.

To formalize the probabilistic setting, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We suppose that for $x \in \mathcal{X}_k$, $A^{(\tau)}(x)$ is a random variable with values in $\mathcal{P}_k$ such that the mapping $(x, \omega) \mapsto A^{(\tau)}(x)(\omega)$ is $\mathcal{M} \otimes \mathcal{F}$-measurable for all τ. We have for all $x \in \mathcal{X}_k$ and $p \in \mathcal{P}_k$: $\pi_p^{(\tau)}(x) = \mathbb{P}[A^{(\tau)}(x) = p]$. In this setting, the distribution $\mu^{k^{(\tau)}}$ is a random variable, as we recall that $\forall p \in \mathcal{P}_k$, $\mu^{k^{(\tau)}}_p = \frac{1}{F_k}\int_{\mathcal{X}_k} \mathbb{1}_{\{A^{(\tau)}(x)=p\}} dm(x)$, and $A^{(\tau)}(x)$ is random. In particular, $\mathbb{E}[\mu^{k^{(\tau)}}_p] = \frac{1}{F_k}\int_{\mathcal{X}_k} \pi_p^{(\tau)}(x)dm(x)$.

Since players are non-cooperative, we consider that players randomize independently. Under this assumption, the distribution $\mu^{(\tau)}$ is almost surely equal to its expectation. Here, non-atomicity is essential.

Proposition 1. *In the non-atomic routing game, if players randomize independently, then for all τ, $\mu^{k^{(\tau)}}$ is a random variable with zero variance.*

This follows from Fubini's theorem. As a result, one can think of the distribution μ as a deterministic variable, although individual players are randomizing.

Definition 3.1 (Online algorithm for routing)**.** *An online algorithm (or update rule) for the routing game, applied by a player $x \in \mathcal{X}_k$, is a deterministic sequence of functions $(h_\tau)_{\tau\in\mathbb{N}}$ such that at iteration τ, h_τ maps the history of losses $(\ell^k(\mu(t)))_{t\le\tau}$ and the current strategy $\pi^{(\tau)}(x)$ to the strategy on the next iteration, $\pi^{(\tau+1)}(x) = h_\tau\left((\ell^k(\mu(t)))_{t\le\tau}, \pi^{(\tau)}(x)\right)$.*

This online learning framework is summarized in Algorithm 1. Here, we assume that, at the end of day τ, a player $x \in \mathcal{X}_k$ observes all the path latencies for her commodity, i.e. $(\ell_p^k(\mu^{(\tau)}))_{p\in\mathcal{P}_k}$. This can be achieved for example by having a central authority publicly re-

port the path latencies at the end of a given day. We note however that the information model could be further restricted such that every player only observes the latency on his/her own path. One appropriate framework to study this problem is that of multi-armed bandit learning, see for example Auer et al. (2002), György et al. (2007), Dani et al. (2008), and Bubeck & Cesa-Bianchi (2012). However, we do not currently consider this extension.

3.2. Discounted regret

The regret is a natural measure of performance of a learning algorithm (Cesa-Bianchi & Lugosi, 2006). In particular, we are interested in online learning algorithms with *sublinear discounted regret*. More precisely, we assume that losses are discounted over time, by a decreasing sequence of factors $(\gamma_\tau)_{\tau\in\mathbb{N}}$. So at iteration τ, a player who chooses path p incurs a loss $\gamma_\tau \ell_p^k(\mu^{(\tau)})$. The sequence $(\gamma_\tau)_\tau$ is assumed to be universal: the discounting is identical across players. This can be justified from an economic perspective if one thinks of discounting as reflecting interest rates.

Assumption 2. *$(\gamma_\tau)_\tau$ is a positive, decreasing, non-summable sequence.*

The idea of discounted regret is common in the online learning literature, and is studied for example by Cesa-Bianchi & Lugosi in (2006). It is worth noting, however, that the sequence is usually assumed to be *increasing*. In our case, discounting the losses by a decreasing sequence can be motivated by the assumption that players value future time less than current time. Given the sequence of discount factors, the discounted regret is defined as follows:

Definition 3.2 (Discounted regret)**.** *Consider a player $x \in \mathcal{X}_k$. Given a sequence of strategies $(\pi^{(\tau)}(x))_\tau$ and a sequence of distributions $(\mu^{(\tau)})_\tau$, the discounted regret of x up to time T is:*

$$R^{(T)}(x) = L^{(T)}(x) - \min_{p\in\mathcal{P}_k} \mathscr{L}_p^{k(T)} \tag{3}$$

where $L^{(T)}(x)$ and $\mathscr{L}_p^{k(T)}$ are, respectively, the expected discounted cumulative loss incurred by x, and the discounted cumulative loss on path $p \in \mathcal{P}_k$:

$$L^{(T)}(x) = \sum_{\tau\le T} \gamma_\tau \sum_{p\in\mathcal{P}_k} \pi_p^{(\tau)}(x)\ell_p^k(\mu^{(\tau)})$$
$$\mathscr{L}_p^{k(T)} = \sum_{\tau\le T} \gamma_\tau \ell_p^k(\mu^{(\tau)})$$

Definition 3.3 (Sublinear discounted regret)**.** *An online learning algorithm for routing $(h_\tau)_\tau$ is said to have sublinear discounted regret if whenever a player x applies the algorithm, for all initial strategies $\pi^{(0)}(x)$ and all sequences $(\mu^{(\tau)})$, $\limsup_{T\to\infty} \frac{1}{\sum_{\tau\le T}\gamma_\tau} R^{(T)}(x) \le 0$.*

An algorithm with sublinear discounted regret performs asymptotically as well as the best constant strategy in hindsight.

3.3. Discounted Hedge algorithm

We now give one example of online learning algorithm with sublinear discounted regret.

Definition 3.4 (Hedge algorithm)**.** *Consider a player $x \in \mathcal{X}_k$. A Hedge algorithm with learning rates $(\gamma_\tau)_\tau$ is an online algorithm $(h_\tau)_\tau$ which satisfies the following update equation:*

$$\pi^{(\tau+1)} \propto \left(\pi_p^{(\tau)} e^{-\gamma_\tau \frac{\ell_p^k(\mu^{(\tau)})}{\rho}}\right)_{p\in\mathcal{P}_k} \tag{4}$$
$$\propto \left(\pi_p^{(0)} e^{-\frac{\mathscr{L}_p^{k(\tau)}}{\rho}}\right)_{p\in\mathcal{P}_k} \tag{5}$$

Next, we give a bound on the discounted regret of the Hedge algorithm, a generalization of Lemma 5.1 in Cesa-Bianchi & Lugosi (2006).

Proposition 2. *If $(\gamma_\tau)_\tau$ is a square-summable sequence satisfying Assumption 2, the Hedge algorithm with learning rates $(\gamma_\tau)_\tau$, applied by a player $x \in \mathcal{X}_k$, has sublinear regret. More precisely, if ρ is a uniform upper bound on the sequence of losses, then*

$$R^{(T)}(x) \le -\rho \log \pi_{\min}^{(0)}(x) + \rho \sum_{\tau\le T} \frac{\gamma_\tau^2}{8}.$$

where $\pi_{\min}^{(0)} = \min_p \pi_p^{(0)}$

Proof. Let $\xi\colon u \in \mathbb{R}_+^{\mathcal{P}_k} \mapsto \log(\sum_{p\in\mathcal{P}_k} \pi_p^{(0)} e^{\frac{u_p}{\rho}})$. By equation (5), we have for all τ:

$$\begin{aligned}
&\xi(\mathscr{L}^{k(\tau)}) - \xi(\mathscr{L}^{k(\tau-1)}) \\
&= \log\left(\frac{\sum_{p\in\mathcal{P}_k} \pi_p^{(0)} e^{-\frac{\mathscr{L}_p^{k(\tau-1)}}{\rho}} e^{-\gamma_\tau \frac{\ell_p^k(\mu^{(\tau)})}{\rho}}}{\sum_{p\in\mathcal{P}_k} \pi_p(0) e^{-\frac{\mathscr{L}_p^{k(\tau-1)}}{\rho}}}\right) \\
&= \log\left(\sum_{p\in\mathcal{P}_k} \pi_p^{(\tau)} e^{-\gamma_\tau \frac{\ell_p^k(\mu^{(\tau)})}{\rho}}\right) \\
&\le -\gamma_\tau \sum_{p\in\mathcal{P}_k} \pi_p^{(\tau)} \frac{\ell_p^k(\mu^{(\tau)})}{\rho} + \frac{\gamma_\tau^2}{8}
\end{aligned}$$

The last inequality follows from Hoeffding's lemma, since $0 \le \ell_p^k(\mu^{(\tau)})/\rho \le 1$.

Summing over τ, we have:

$$\xi(\mathscr{L}_p^{k^{(T)}}) \leq -\frac{L^{(T)}(x)}{\rho} + \sum_{\tau \leq T} \frac{\gamma_\tau^2}{8}$$

As log is increasing, $\xi(\mathscr{L}_p^{k^{(T)}}) \geq \log(\pi_p^{(0)}) + \mathscr{L}_p^{k^{(\tau)}}/\rho$ for all $p \in \mathcal{P}_k$. Rearranging, we have:

$$L^{(T)}(x) - \mathscr{L}_p^{k^{(T)}} \leq -\rho \log \pi_p^{(0)} + \rho \sum_{\tau \leq T} \frac{\gamma_\tau^2}{8}$$

we conclude by maximizing both sides over $p \in \mathcal{P}_k$. □

Given the previous Proposition, discounting losses can be interpreted, in the case of the Hedge algorithm, as using a decreasing sequence of *learning rates* $(\gamma_\tau)_\tau$.

3.4. Population regret

We define the discounted regret for population $\mathcal{X}_k$ by integrating the individual regrets of players:

$$R^{k^{(T)}} = \frac{1}{F_k} \int_{\mathcal{X}_k} R^{(T)}(x) dm(x)$$

If we define the average cumulative loss of population $\mathcal{X}_k$ to be $L^{k^{(T)}} = \frac{1}{F_k} \int_{\mathcal{X}_k} L^{(T)}(x) dm(x) = \sum_{\tau \leq T} \gamma_\tau \left\langle \mu^{k^{(\tau)}}, \ell^k(\mu^{(\tau)}) \right\rangle$, then we also have $R^{k^{(T)}} = L^{k^{(T)}} - \min_{p \in \mathcal{P}_k} \mathscr{L}_p^{k^{(T)}}$. As a consequence of this definition, if all players in $\mathcal{X}_k$ apply algorithms with sublinear discounted regret, the population-wide regret is also sublinear, that is, $\limsup_{T \to \infty} \frac{1}{\sum_{\tau \leq T} \gamma_\tau} R^{k^{(T)}} \leq 0$.

4. Convergence to Nash equilibria

4.1. Convergence on almost all days

We give a first convergence result. For $\mu \in \Delta$, let $d(\mu, \mathcal{N}) = \inf_{\nu \in \mathcal{N}} \|\mu - \nu\|$ where $\|\cdot\|$ is the Euclidean distance on $\mathbb{R}^{\mathcal{P}}$. We say that a sequence $(\mu^{(\tau)})_\tau$ converges to the set $\mathcal{N}$ if $d(\mu^{(\tau)}, \mathcal{N}) \to 0$.

Proposition 3 (Statistical convergence to Nash equilibria)**.** *Consider a routing game with population dynamics such that the population regret is sublinear, and let $(\mu^{(\tau)})_\tau$ be the sequence of path distributions. Then there exists a subsequence $(\mu^{(\tau)})_{\tau \in \mathcal{T}}$ which converges to $\mathcal{N}$, defined on a subset $\mathcal{T} \subset \mathbb{N}$ of density one, that is,* $\lim_{T \to \infty} \frac{\sum_{\tau \in \mathcal{T} : \tau \leq T} \gamma_\tau}{\sum_{\tau \in \mathbb{N} : \tau \leq T} \gamma_\tau} = 1$.

In other words, the strategies converge *on almost all days* if the population regret is sublinear. This is a limit case in Theorem 5.1 in (Blum et al., 2006). We present a different proof which uses the Rosenthal potential function, and which holds even if the latency functions are not Lipschitz continuous. We first need the following technical Lemma.

Lemma 1. *Let $(\gamma_\tau)_{\tau \in \mathbb{N}}$ be a non-summable sequence of positive weights. If a real sequence $(u^{(\tau)})_{\tau \in \mathbb{N}}$ converges absolutely to u in the sense of Cesàro means w.r.t. $(\gamma_\tau)_\tau$, that is* $\lim_{T \to \infty} \frac{\sum_{\tau \leq T} \gamma_\tau |u^{(\tau)} - u|}{\sum_{\tau \leq T} \gamma_\tau} = 0$*, then there exists a subset of indexes $\mathcal{T}$ of density one such that the subsequence $(u^{(\tau)})_{\tau \in \mathcal{T}}$ converges to u.*

Proof of Proposition 3. Let $\mu^\star \in \mathcal{N}$ be a Nash equilibrium, i.e. $\mu^\star \in \arg\min_{\mu \in \Delta} V$ (see Theorem 1). Then by convexity of V and Equation (2),

$$\begin{aligned} V(\mu^{(\tau)}) - V_{\mathcal{N}} &\leq \left\langle \nabla V(\mu^{(\tau)}), \mu^{(\tau)} - \mu^\star \right\rangle \\ &\leq \sum_{k=1}^{K} F_k \left\langle \ell^k(\mu^{(\tau)}), \mu^{k^{(\tau)}} - \mu^{\star k} \right\rangle \end{aligned}$$

here we use $\langle \cdot, \cdot \rangle$ to denote the inner product on $\mathbb{R}^{\mathcal{P}}$. Then, taking the weighted sum up to time T,

$$\begin{aligned} &\frac{\sum_{\tau \leq T} \gamma_\tau (V(\mu^{(\tau)}) - V_{\mathcal{N}})}{\sum_{\tau \leq T} \gamma_\tau} \\ &\leq \sum_{k=1}^{K} F_k \frac{\sum_{\tau \leq T} \gamma_\tau \left\langle \mu^{k^{(\tau)}}, \ell^k(\mu^{(\tau)}) \right\rangle - \left\langle \mu^{\star k}, \mathscr{L}^{k^{(T)}} \right\rangle}{\sum_{\tau \leq T} \gamma_\tau} \\ &\leq \sum_{k=1}^{K} F_k \frac{R^{k^{(T)}}}{\sum_{\tau \leq T} \gamma_\tau} \end{aligned}$$

where the last inequality follows from the fact that $\left\langle \mu^\star, \mathscr{L}^{k^{(T)}} \right\rangle \geq \min_p \mathscr{L}_p^{k^{(T)}}$. Since, for all τ, $V(\mu^{(\tau)}) - V_{\mathcal{N}} \geq 0$, and for all k, $\limsup_{T \to \infty} \frac{1}{\sum_{\tau \leq T} \gamma_\tau} R^{k^{(T)}} \leq 0$ by assumption, we have $(V(\mu^{(\tau)}))_\tau$ converges absolutely to $V_{\mathcal{N}}$ in the sense of Cesàro means w.r.t. $(\gamma_\tau)_\tau$. Thus by Lemma 1, there exists a dense subset of indexes $\mathcal{T}$ such that $(V(\mu^{(\tau)}))_{\tau \in \mathcal{T}}$ converges to $V_{\mathcal{N}}$, and by continuity of V and compactness of Δ, the subsequence $(\mu^{(\tau)})_{\tau \in \mathcal{T}}$ converges to $\mathcal{N}$. □

In order to show strong convergence for a class of online algorithms with sublinear discounted regret, we first study the continuous-time replicator dynamics, which can be motivated as a continuous-time limit of the Hedge algorithm, as discussed next.

4.2. Continuous-time dynamics

We consider the discounted Hedge algorithm with a vanishing sequence of learning rates (γ_τ), acting on the sequence of population strategies $(\mu^{(\tau)})_{\tau \in \mathbb{N}}$.

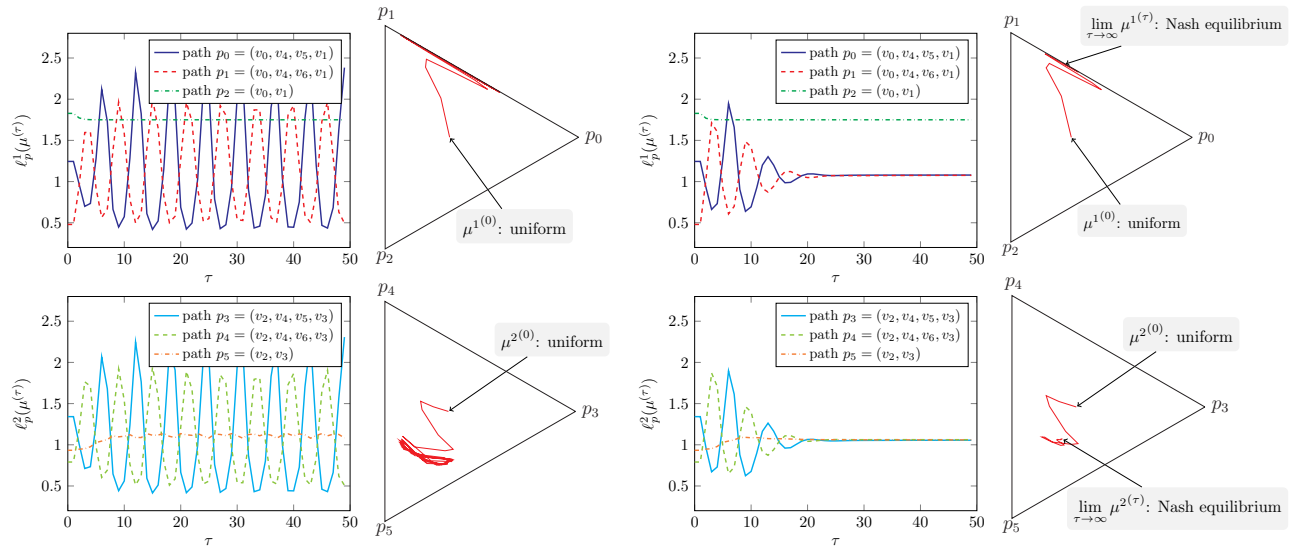

Figure 2. Example of a routing game payed on the example network of Figure 1. Latency functions are taken to be quadratic increasing, and generated randomly. The population strategies $(\mu^{k(\tau)})_\tau$ obey the Hedge algorithm. The figures show the trajectories in the simplex $\Delta^{\mathcal{P}_k}$, and the resulting path latencies $(\ell_p^k(\mu^{(\tau)}))$ for population $\mathcal{X}_1$ (top) and $\mathcal{X}_2$ (bottom). With a constant learning rate $\gamma = 0.7$, $(\mu^{k(\tau)})_\tau$ does not converge (left). With a harmonic sequence of learning rates, $\gamma_\tau = \frac{1}{1+\tau/10}$, $(\mu^{k(\tau)})_\tau$ converges to the set of Nash equilibria.

Let us imagine an underlying continuous time $T \in \mathbb{R}_+$, and set $\boldsymbol{\mu}(T_\tau) = \mu^{(\tau)}$, where T_τ is the time at which the τ-th update happens. Now choosing the update times to be $T_\tau = \sum_{t=1}^{\tau} \gamma_t$, we can write, $\forall p \in \mathcal{P}_k$

$$\begin{aligned}
\boldsymbol{\mu}_p^k(T_{\tau+1}) &= \mu_p^{k(\tau+1)} \\
&= \mu_p^{k(\tau)} \frac{e^{-\gamma_\tau \ell_p^k(\mu^{(\tau)})/\rho}}{\sum_{p' \in \mathcal{P}_k} \mu_{p'}^{k(\tau)} e^{-\gamma_\tau \ell_{p'}^k(\mu^{(\tau)})/\rho}} \\
&= \mu_p^{k(\tau)} \frac{1 - \gamma_\tau \ell_p^k(\mu^{(\tau)})/\rho + o(\gamma_\tau)}{1 - \gamma_\tau \sum_{p' \in \mathcal{P}_k} \mu_{p'}^{k(\tau)} \ell_{p'}^k(\mu^{(\tau)})/\rho + o(\gamma_\tau)} \\
&= \boldsymbol{\mu}_p^k(T_\tau) \left[1 + \gamma_\tau \frac{\langle \mu^{(\tau)}, \ell^k(\mu^{(\tau)}) \rangle - \ell_p^k(\mu^{(\tau)})}{\rho} \right] + o(\gamma_\tau)
\end{aligned}$$

Thus,

$$\frac{\boldsymbol{\mu}_p^k(T_\tau + \gamma_\tau) - \boldsymbol{\mu}_p^k(T_\tau)}{\gamma_\tau} = \boldsymbol{\mu}_p^k(T_\tau) \frac{\langle \boldsymbol{\mu}^k(T_\tau), \ell^k(\boldsymbol{\mu}(T_\tau)) \rangle - \ell_p^k(\boldsymbol{\mu}(T_\tau))}{\rho} + o(1)$$

taking the limit of the above equation as $\gamma_\tau \to 0$, we obtain the following ODE

$$\begin{cases} \boldsymbol{\mu}(0) \in \mathring{\Delta} \\ \frac{d\boldsymbol{\mu}(t)}{dt} = G(\boldsymbol{\mu}(t), \ell(\boldsymbol{\mu}(t))) \end{cases} \tag{6}$$

where $\forall k$ and $\forall p \in \mathcal{P}_k$

$$G_p^k(\boldsymbol{\mu}, \ell) = \boldsymbol{\mu}_p^k \frac{\langle \boldsymbol{\mu}^k, \ell^k \rangle - \ell_p^k}{\rho} \tag{7}$$

Here, $\mathring{\Delta} = \{\mu \in \Delta \colon \forall p \in \mathcal{P}, \mu_p > 0\}$ is the relative interior of Δ. Starting in $\mathring{\Delta}$ guarantees that $\boldsymbol{\mu}(t)$ remains in $\mathring{\Delta}$ for all t. In this derivation, the discount factors γ_τ are interpreted as discrete time steps. The dynamics described by this ODE, called the replicator dynamics (Fischer & Vöcking, 2004), has been studied extensively. One can observe in particular that the set $\mathrm{R}\mathcal{N}$ of restricted Nash equilibria (Definition 2.2) is exactly the set of stationary points for the ODE.

4.3. Replicator updates

By discretizing the replicator dynamics, we obtain a multiplicative update rule we call REP for Replicator, which has desirable properties which we prove next.

Definition 4.1 (REP algorithm)**.** *The replicator (REP) algorithm with rates $(\gamma_\tau)_\tau$, $\gamma_\tau \leq 1$, applied by $x \in \mathcal{X}_k$, is an online algorithm for routing given by the following update equation*

$$\pi_p^{(\tau+1)} - \pi_p^{(\tau)} = \gamma_\tau G_p^k(\pi^{(\tau)}, \ell(\mu^{(\tau)})) \tag{8}$$

We note that summing this update equation over $p \in \mathcal{P}_k$ yields $\sum_{p \in \mathcal{P}_k} (\pi_p^{(\tau+1)} - \pi_p^{(\tau)}) = 0$, thus π remains

in $\Delta^{\mathcal{P}_k}$ as long as $\gamma_\tau \leq 1$. We now show that the REP update rule guarantees a sublinear discounted regret. To see this, we need the following regret bound on multiplicative-weights updates with signed losses.

Lemma 2. *Consider an online learning setting with signed losses $s_p^{(\tau)} \in [-1,1]$ and discount factors $\gamma_\tau \leq 1/2$ satisfying Assumption 2. Then the discounted multiplicative-weights algorithm defined by*

$$\pi^{(\tau+1)} \propto \left(\pi_p^{(\tau)}(1-\gamma_\tau s_p^{(\tau)})\right)_p \tag{9}$$

guarantees that for any p,

$$\sum_{\tau\leq T}\gamma_\tau\left(\left\langle \pi^{(\tau)}, s^{(\tau)}\right\rangle - s_p^{(\tau)}\right) \leq -\log \pi_{\min}^{(0)} + \sum_{\tau\leq T}\gamma_\tau^2$$

This Lemma is a straightforward extension of Theorem 2.1 in (Arora et al., 2012) to the discounted case.

Proposition 4. *If $(\gamma_\tau)_\tau$ is a square-summable sequence of discount factors satisfying Assumption 2 and such that $\gamma_\tau \leq 1/2$ for all τ, the (REP) update rule with rates $(\gamma_\tau)_\tau$ has sublinear discounted regret.*

Proof. Let $r_p^{(\tau)} = \left\langle \pi^{(\tau)}, \ell^k(\mu^{(\tau)})\right\rangle - \ell_p^k(\mu^{(\tau)}) \in [-\rho,\rho]$ be the *instantaneous regret* of the player. Then the REP update can be viewed as a multiplicative-weights algorithm with update rule (9), signed losses $s_p^{(\tau)} = -r_p^{k(\tau)}/\rho \in [-1,1]$, and discount factors γ_τ. Observing that $\left\langle r^{k(\tau)}, \pi^{(\tau)}\right\rangle = 0$, we have by Lemma 2:

$$\frac{1}{\rho}\sum_{\tau\leq T}\gamma_\tau r_p^{k(\tau)} \leq -\log \pi_{\min}(0) + \sum_{\tau\leq T}\gamma_\tau^2$$

Rearranging and taking the maximum over $p \in \mathcal{P}_k$, we obtain the following bound on the discounted regret

$$R^{(T)}(x) \leq -\rho \log \pi_{\min}^{(0)} + \rho\sum_{\tau\leq T}\gamma_\tau^2$$

which shows $\limsup_{T\to\infty}\frac{1}{\sum_{\tau\leq T}\gamma_\tau}R^{(T)}(x) \leq 0$. □

4.4. Approximate Replicator algorithms

Definition 4.2 (AREP algorithm)**.** *An online algorithm for routing, applied by $x \in \mathcal{X}_k$, is said to be an approximate replicator algorithm (AREP) if its update equation can be written as*

$$\pi_p^{(\tau+1)} - \pi_p^{(\tau)} = \gamma_\tau(G_p^k(\pi^{(\tau)}, \ell(\mu^{(\tau)})) + U_p^{(\tau+1)}) \tag{10}$$

where $(U^{(\tau)})_{\tau\geq 1}$ is a bounded sequence of stochastic perturbations with values in $\mathbb{R}^{\mathcal{P}_k}$, and which satisfies the following condition: for all $T > 0$,

$$\lim_{\tau_1\to\infty}\max_{\tau_2:\ \sum_{\tau=\tau_1}^{\tau_2}\gamma_\tau < T}\left\|\sum_{\tau=\tau_1}^{\tau_2}\gamma_\tau U^{(\tau+1)}\right\| = 0 \tag{11}$$

Condition (11) corresponds to the first hypothesis of Proposition 4.1 in (Benaïm, 1999), which we will use in the proof of the main convergence theorem. It bounds the cumulative perturbation over a given time interval T. Intuitively, this condition will ensure that the trajectories of a discrete AREP algorithm are asymptotically close to the trajectories of the continuous-time replicator dynamics.

Note that the REP update rule is an AREP algorithm with zero perturbation. By allowing perturbations, we extend the class of algorithms for which we can show convergence. In particular, we show that the discounted Hedge algorithm is in this class.

Proposition 5. *The Hedge algorithm with learning rates $(\gamma_\tau)_\tau$ satisfying Assumption 2 with $\sum_\tau \gamma_\tau^2 < \infty$ is an AREP algorithm.*

Proof. Let $(\pi^{(\tau)})_{\tau\in\mathbb{N}}$ be the sequence of player strategies, and $(\mu^{(\tau)})_\tau$ be any sequence of population distributions. By definition of the Hedge algorithm,

$$\pi_p^{(\tau+1)} = \pi_p^{(\tau)} e^{-\gamma_\tau\frac{\ell_p^k(\mu^{(\tau)})}{\rho}} \Big/ \sum_{p'\in\mathcal{P}_k}\pi_{p'}^{(\tau)} e^{-\gamma_\tau\frac{\ell_{p'}^k(\mu^{(\tau)})}{\rho}}$$

which we can write in the form of equation (10), with

$$\begin{aligned}
U_p^{(\tau+1)} &= \frac{\pi_p^{(\tau)}}{\gamma_\tau}\Big[e^{-\gamma_\tau\frac{\ell_p^k(\mu^{(\tau)})-\tilde{\ell}^{k(\tau)}}{\rho}} + \gamma_\tau\frac{\ell_p^k(\mu^{(\tau)})-\tilde{\ell}^{k(\tau)}}{\rho} \\
&\quad - 1\Big] + \pi_p^{(\tau)}\frac{\tilde{\ell}^{k(\tau)}-\bar{\ell}^{k(\tau)}}{\rho} \\
\tilde{\ell}^{k(\tau)} &= -\frac{\rho}{\gamma_\tau}\log\sum_{p'\in\mathcal{P}_k}\pi_{p'}^{(\tau)} e^{-\gamma_\tau \ell_{p'}^k(\mu^{(\tau)})/\rho} \\
\bar{\ell}^{k(\tau)} &= \left\langle \pi^{(\tau)}, \ell^k(\mu^{(\tau)})\right\rangle
\end{aligned}$$

Letting $\theta(x) = e^x - x - 1$, we have for all $p \in \mathcal{P}_k$:

$$\begin{aligned}
U_p^{(\tau+1)} &= \frac{\pi_p^{(\tau)}}{\gamma_\tau}\theta\left(-\gamma_\tau\frac{\ell_p(\mu^{(\tau)})-\tilde{\ell}^{k(\tau)}}{\rho}\right) \\
&\quad + \frac{\pi_p^{(\tau)}}{\rho}\left(\tilde{\ell}^{k(\tau)} - \bar{\ell}^{k(\tau)}\right)
\end{aligned}$$

The first term is a $O(\gamma_\tau)$ as $\theta(x) \sim_0 x^2/2$. To bound the second term, we have by concavity of the logarithm

$$\tilde{\ell}^{k(\tau)} \leq \sum_{p'\in\mathcal{P}_k}\pi_{p'}^{(\tau)}\ell_{p'}^k(\mu^{(\tau)}) = \bar{\ell}^{k(\tau)}$$

And by Hoeffding's lemma,

$$\log\sum_{p'\in\mathcal{P}_k}\pi_{p'}e^{-\gamma_\tau\frac{\ell_{p'}(\mu^{(\tau)})}{\rho}} \leq -\gamma_\tau\sum_{p'\in\mathcal{P}_k}\pi_{p'}^{(\tau)}\frac{\ell_{p'}(\mu^{(\tau)})}{\rho} + \frac{\gamma_\tau^2}{8}$$

Rearranging, we have $0 \leq \bar{\ell}^{k(\tau)} - \tilde{\ell}^{k(\tau)} \leq \frac{\rho\gamma_\tau}{8}$, therefore $U_p^{(\tau+1)} = O(\gamma_\tau)$, and $\|\sum_{\tau=\tau_1}^{\tau_2}\gamma_\tau U^{(\tau+1)}\| = O(\sum_{\tau=\tau_1}^{\tau_2}\gamma_t^2)$. Condition (11) is thus verified. □

4.5. Strong convergence to Nash equilibria

Theorem 2. *If for all k, the population strategies $(\mu^{k(\tau)})_\tau$ satisfy an AREP algorithm with sublinear regret, then the sequence $(\mu^{(\tau)})$ converges to the set of Nash equilibria.*

Proof. The proof proceeds in two steps: first, we use results from (Benaïm, 1999) to prove that the sequence of potentials $V(\mu^{(\tau)})$ converges. Then, using convergence on most days given in Proposition 3, we conclude that $V(\mu^{(\tau)})$ converges necessarily to the minimum $V_{\mathcal{N}}$, which proves that $(\mu^{(\tau)})$ converges to $\mathcal{N}$ by continuity of V on the compact Δ. First, we recall the definition of Lyapunov function.

Definition 4.3 (Lyapunov function)**.** *Let $\Gamma \subset \Delta$ be a compact invariant set for the replicator ODE (6). A continuous non-negative function $V : \Delta \to \mathbb{R}_+$ is a Lyapunov function for Γ if $\frac{d}{dt}V(\boldsymbol{\mu}(t)) = \langle \nabla V(\boldsymbol{\mu}(t)), G(\boldsymbol{\mu}(t), \ell(\boldsymbol{\mu}(t)))\rangle < 0$ for all $\boldsymbol{\mu}(t) \notin \Gamma$.*

Lemma 3 (Convergence of potentials under AREP algorithms)**.** *Let Γ be a compact invariant set for the replicator ODE (6), V a Lyapunov function for Γ, and assume $V(\Gamma)$ has empty interior. Assume that the sequence of distributions $(\mu^{(\tau)})_{\tau\in\mathbb{N}}$ obeys an AREP update rule. Then the sequence of potentials $(V(\mu^{(\tau)}))_\tau$ converges.*

This follows from Theorem 5.7 and Proposition 4.1 in (Benaïm, 1999). Here, condition (11) is essential.

Next, we show that the Rosenthal potential function V is a Lyapunov function for the invariant set $\mathrm{R}\mathcal{N}$ of restricted Nash equilibria. From equation (2) and the definition of G,

$$\begin{aligned}
&\langle \nabla V(\boldsymbol{\mu}(t)), G(\boldsymbol{\mu}(t), \ell(\boldsymbol{\mu}(t)))\rangle \\
&= \sum_k F_k \sum_{p\in\mathcal{P}_k} \ell_p^k(\boldsymbol{\mu}(t))\boldsymbol{\mu}_p(t)\left(\left\langle \boldsymbol{\mu}(t), \ell_p^k(\boldsymbol{\mu}(t))\right\rangle - \ell_p^k(\boldsymbol{\mu}(t))\right) \\
&= \sum_k F_k \left[\left(\sum_{p\in\mathcal{P}_k} \boldsymbol{\mu}_p(t)\ell_p^k(\boldsymbol{\mu}(t))\right)^2 - \sum_{p\in\mathcal{P}} \boldsymbol{\mu}_p(t)\ell_p^k(\boldsymbol{\mu}(t))^2\right]
\end{aligned}$$

which is less than or equal to 0 by Jensen's inequality, with equality if and only if $\mu \in \mathrm{R}\mathcal{N}$. Therefore V is a Lyapunov function for $\mathrm{R}\mathcal{N}$. And since $V(\mathrm{R}\mathcal{N})$ is a finite set by Remark 1, it has empty interior relatively to $\mathbb{R}$, and we can apply Lemma 3, and conclude that the sequence of potentials $(V(\mu^{(\tau)}))_{\tau\in\mathbb{N}}$ converges. It remains to show that its limit is $V_{\mathcal{N}}$.

Since the AREP algorithm is assumed to have sublinear discounted regret, we can apply Proposition 3: there exists a dense subsequence $(\mu^{(\tau)})_{\tau\in\mathcal{T}}$ which converges to $\mathcal{N}$. The corresponding subsequence of potentials $(V(\mu^{(\tau)}))_{\tau\in\mathcal{T}}$ converges to $V_{\mathcal{N}}$ by continuity of V, and by uniqueness of the limit, we must have $\lim_\tau V(\mu^{(\tau)}) = V_{\mathcal{N}}$. This concludes the proof. □

Corollary 1. *If $(\gamma_\tau)_\tau$ be a square-summable sequence bounded by $1/2$ satisfying Assumption 2, and $(\mu_\tau)_\tau$ obeys the REP update rule with rates (γ_τ), then $(\mu_\tau)_\tau$ converges to the set of Nash equilibria.*

Proof. Under these assumptions, the REP algorithm has sublinear discounted regret by Proposition 4. It is also an AREP algorithm (with zero perturbations) so we can apply Theorem 2. □

Corollary 2. *If $(\gamma_\tau)_\tau$ is a square-summable sequence satisfying Assumption 2, and $(\mu_\tau)_\tau$ obeys the discounted Hedge algorithm with rates (γ_τ), then $(\mu_\tau)_\tau$ converges to the set of Nash equilibria.*

Proof. By Proposition 2 and Proposition 5, the Hedge algorithm with rates γ_τ is an AREP algorithm with sublinear discounted regret, and we can apply Theorem 2. □

Figure 2 shows an example of discounted Hedge algorithm with a non-summable, square-summable sequence of learning rates. The resulting strategies converge to the set of Nash equilibria.

5. Conclusion

In order to obtain strong convergence guarantees of online learning algorithms applied to routing games, we consider a model in which losses are discounted. We studied a continuous-time limit of the Hedge algorithm. This motivated the introduction of a class of no-regret learning algorithms, called AREP, which can be viewed as approximations of the replicator dynamics. Using results from the theory of stochastic approximation, we showed that under this class, $(\mu^{(\tau)})$ is guaranteed to converge to the set of Nash equilibria.

These results assume a universal sequence $(\gamma_\tau)_\tau$ of discounts; thus a natural question is whether convergence still holds if this assumption is relaxed. Another open question is whether the learning algorithm is robust to observation noise: if latency observations are noisy, with bounded noise, can one guarantee convergence if the bound is small enough?

References

Arora, Sanjeev, Hazan, Elad, and Kale, Satyen. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

Auer, Peter, Cesa-Bianchi, Nicolò, and Fischer, Paul. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2-3):235–256, May 2002.

Beckmann, Martin J, McGuire, Charles B, and Winsten, Christopher B. Studies in the economics of transportation. 1955.

Benaïm, Michel. Dynamics of stochastic approximation algorithms. In *Séminaire de probabilités XXXIII*, pp. 1–68. Springer, 1999.

Blum, Avrim, Even-Dar, Eyal, and Ligett, Katrina. Routing without regret: on convergence to nash equilibria of regret-minimizing algorithms in routing games. In *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, PODC '06, pp. 45–52, New York, NY, USA, 2006. ACM.

Bubeck, Sébastien and Cesa-Bianchi, Nicolò. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Cesa-Bianchi, Nicolò and Lugosi, Gábor. *Prediction, learning, and games*. Cambridge University Press, 2006.

Dafermos, Stella C and Sparrow, Frederick T. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards, Series B*, 73(2):91–118, 1969.

Dani, Varsha, Hayes, Thomas, and Kakade, Sham. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems 20*, pp. 345–352, Cambridge, MA, 2008. MIT Press.

Fischer, Simon and Vöcking, Berthold. On the evolution of selfish routing. In *Algorithms–ESA 2004*, pp. 323–334. Springer, 2004.

Fischer, Simon, Räcke, Harald, and Vöcking, Berthold. Fast convergence to wardrop equilibria by adaptive sampling methods. *SIAM Journal on Computing*, 39(8):3700–3735, 2010.

Freund, Yoav and Schapire, Robert E. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1):79–103, 1999.

György, András, Linder, Tamás, Lugosi, Gábor, and Ottucsák, György. The on-line shortest path problem under partial monitoring. *Journal of Machine Learning Research*, 8:2369–2403, December 2007.

Littlestone, Nick and Warmuth, Manfred K. The weighted majority algorithm. In *Foundations of Computer Science, 1989., 30th Annual Symposium on*, pp. 256–261. IEEE, 1989.

Monderer, Dov and Shapley, Lloyd S. Fictitious play property for games with identical interests. *journal of economic theory*, 68(1):258–265, 1996.

Papadimitriou, Christos H. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and system Sciences*, 48(3):498–532, 1994.

Rosenthal, Robert W. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

Roughgarden, T. Routing games. In *Algorithmic game theory*, chapter 18, pp. 461–486. Cambridge University Press, 2007.

Sandholm, William H. Potential games with continuous player sets. *Journal of Economic Theory*, 97(1): 81–108, 2001.

Wardrop, John Glen. Some theoretical aspects of road traffic research. In *ICE Proceedings: Engineering Divisions*, volume 1, pp. 325–362. Thomas Telford, 1952.

Weibull, Jörgen W. *Evolutionary game theory*. MIT press, 1997.