Bayesian Learning with Adaptive Load Allocation Strategies

Manxi Wu MANXIWU@MIT.EDU
Institute of Data, Systems, and Society, MIT, Cambridge, MA, USA.

Saurabh Amin AMINS@MIT.EDU
Laboratory of Information and Decision Systems, MIT, Cambridge, MA, USA.

Asuman E. Ozdaglar ASUMAN@MIT.EDU
Laboratory of Information and Decision Systems, MIT, Cambridge, MA, USA.

Editors: A. Bayen, A. Jadbabaie, G. J. Pappas, P. Parrilo, B. Recht, C. Tomlin, M.Zeilinger

Abstract

We study a Bayesian learning dynamics induced by agents who repeatedly allocate loads on a set of resources based on their belief of an unknown parameter that affects the cost distributions of resources. In each step, *belief update* is performed according to Bayes' rule using the agents' current load and a realization of costs on resources that they utilized. Then, agents choose a new load using an adaptive *strategy update* rule that accounts for their preferred allocation based on the updated belief. We prove that beliefs and loads generated by this learning dynamics converge almost surely. The convergent belief accurately estimates cost distributions of resources that are utilized by the convergent load. We establish conditions on the initial load and strategy updates under which the cost estimation is accurate on *all* resources. These results apply to Bayesian learning in congestion games with unknown latency functions. Particularly, we provide conditions under which the load converges to an equilibrium or socially optimal load with complete information of cost parameter. We also design an adaptive tolling mechanism that eventually induces the socially optimal outcome.

Keywords: Bayesian learning, Congestion games, Adaptive pricing mechanisms

1. Introduction

1.1. Learning dynamics

We consider a situation in which one or more agents (players) allocate non-negative loads on a finite set of resources E. The aggregate effect of agents' allocation strategy is captured by a load vector, denoted $x = (x_e)_{e\in E}$, where x_e is the total load (or level of utilization) on resource e. The set of feasible load vectors is a convex continuous set $X \subseteq \mathbb{R}^{|E|}_{\geq 0}$. The cost of using resource e, denoted y_e, is a continuous random variable, which may be correlated with the costs of other resources. We denote the vector of costs for all resources as $y = (y_e)_{e\in E}$, and the vector of costs for the resources that are utilized by load vector x as $\hat{y} = (\hat{y}_e)_{e\in \widehat{E}}$. Here $\widehat{E} \triangleq \{E|x_e > 0\}$ denotes the set of resources with non-zero load. We refer y as the full cost vector and $\hat{y}$ as the observed cost vector.

In our setup, the joint probability distribution of cost vector depends on a (scalar or vector) parameter s, which takes values in a finite set S. The true parameter governing the cost distribution, denoted $s^* \in S$, is *unknown*. For a given load vector x and cost parameter s, the probability density function of full cost vector y is $\phi^s(y|x)$. The probability density function of the observed cost vector $\hat{y}$, denoted $\widehat{\phi}^s(\hat{y}|x)$, is the marginal of $\phi^s(y|x)$ on the set $\widehat{E}$. We assume that the following holds:

(A1) *The probability density function $\phi^s(y|x)$ is continuous in x for all $s \in S$.*

We focus on learning of the unknown cost parameter s under the discrete-time stochastic dynamics governing the evolution of the belief distribution $\theta^k \in \Delta(S)$ and the load vector $x^k \in X$ for the time steps $k = 1, 2, \ldots$. The state of learning dynamics in step k is defined by (θ^k, x^k). In each step k, the set of utilized resources is $\widehat{E}^k \triangleq \{E | x_e^k > 0\}$ and observed cost vector $\hat{y}^k \triangleq \left(\hat{y}_e^k\right)_{e \in \widehat{E}^k}$ is realized according to the probability density function $\widehat{\phi}^s(\hat{y}^k | x^k)$, which is the marginal of $\phi^s(y|x^k)$ on the set $\widehat{E}^k$. Specifically, the dynamics of $(\theta^k, x^k)_{k=1}^{\infty}$ is described by the following update rule:

$$\theta^{k+1}(s) = \frac{\theta^k(s)\widehat{\phi}^s(\hat{y}^k|x^k)}{\sum_{s' \in S} \theta^k(s')\widehat{\phi}^{s'}(\hat{y}^k|x^k)}, \quad \forall s \in S, \tag{1a}$$

$$x^{k+1} = (1 - a^k)x^k + a^k g\left(\theta^{k+1}\right). \tag{1b}$$

That is, the belief θ^{k+1} is obtained by Bayesian update from θ^k based on the randomly realized cost vector $\hat{y}^k$ and current load vector x^k. The new load vector x^{k+1} is obtained by taking a linear combination of x^k and $g(\theta^k) : \Delta(S) \to X$, where scalar a^k is the step size of update.

From a game-theoretic viewpoint, the function $g(\cdot)$ captures the aggregate effect of the strategic choices made by the agents in allocating the load to various resources, given the updated (common) belief of the cost parameter. In other words, the function g captures the outcome of agents' "preferred" allocation strategy based on the updated belief. The exact form of $g(\theta)$ depends on how the unknown cost parameter affects the agents' individual payoffs. The step size a^k in (1b) determines the relative weight of the previous load vector x^k and the allocation $g(\theta^{k+1})$ based on agents' preference. In fact, $\left(a^k\right)_{k=1}^{\infty}$ may be exogenously given, or endogenously determined based on states; for e.g. agents may adaptively choose a^k based on the θ^{k+1} and $g(\theta^{k+1})$. We refer (1a) (resp. (1b)) as the *belief update* (resp. *strategy update*) of the learning dynamics.

We make the following (mild) assumptions:

(A2) *$\theta^1(s) > 0$ for all $s \in S$.*

(A3) *$a^k \in (0, 1]$ for all k and $\prod_{k=1}^{\infty}\left(1 - a^k\right) = 0$.*

(A4) *$g(\theta)$ is continuous in θ.*

Note that **(A2)** ensures that the belief in any step does not exclude the true parameter s^*, i.e., $\theta^k(s^*) > 0$ for all $k \geq 1$ with probability (w.p.) 1. In **(A3)**, $a^k \in (0, 1]$ ensures that for any x^k and $g(\theta^k)$ in the convex set X, the updated load vector $x^{k+1} \in X$. We do not assume that the step sizes $(a^k)_{k=1}^{\infty}$ are constant or diminishing as k becomes large, but just require that $\prod_{k=1}^{\infty}\left(1 - a^k\right) = 0$. The assumption of constant or diminishing step sizes, which is typical in stochastic recursive methods Tsitsiklis (1994), Borkar and Meyn (2000), may be limiting when it comes to modeling how agents adaptively change their strategy based on updated beliefs. For e.g., based on realized cost vector $\hat{y}^k$ – which affects the updated belief θ^{k+1} – the agents may choose $g(\theta^{k+1})$ as their strategy in some steps ($a^k = 1$), and put very low weight on $g(\theta^{k+1})$ in other steps (a^k close to 0). On the other hand, the assumption $\prod_{k=1}^{\infty}\left(1 - a^k\right) = 0$ only imposes a mild restriction on a^k when it indeed converges to zero. In particular, it is satisfied as long as a^k does not converge to zero asymptotically faster than the rate $1 - e^{-\frac{1}{k}}$. Besides, it trivially holds when a^k is lower-bounded by a small positive number for all k.

1.2. Our contributions and related literature

In this paper, we analyze (1) to address the problem of Bayesian learning of the unknown parameter, based on step-dependent cost observations on utilized resources and load allocations generated by the update rule. In Section 2, we first show that the learning dynamics converges to a fixed point

almost surely. We also establish an exponential convergence rate of the belief of the unknown parameter. Secondly, when the initial load vector utilizes all resources and $a^k \in (0,1)$ for all k, the convergent belief forms an accurate estimation of the observed cost distribution with the convergent load vector on all resources, including the ones not utilized at fixed point. However, in general, one can only guarantee that the belief forms an accurate estimation on utilized resources at fixed point.

Our analysis approach draws from the fundamental ideas from learning in games Fudenberg and Kreps (1995), Monderer and Shapley (1996), Shamma and Arslan (2005), Cominetti et al. (2010), Krichene et al. (2014), and learning in control systems Tsitsiklis (1994), Borkar and Meyn (2000), Recht (2019). The distinguishing feature of our model is that it captures the dynamic interaction between *(i)* information aggregation via Bayesian belief updates, and *(ii)* adaptive allocation via strategy updates. This interaction is key in the study of statistical learning in strategic environments. Related literature on information aggregation via Bayesian learning includes Gale and Kariv (2003), Acemoglu et al. (2011), and Jadbabaie et al. (2013).

In Section 3, we extend our results to study Bayesian learning in congestion games, where an unknown parameter affects the latency (or average cost) of congestible resources. We show that under certain assumptions on the latency functions, the step-sizes, and full exploration at initial state, the fixed point corresponds to the complete information equilibrium (resp. socially optimal) load vector, when the $g(\cdot)$ function for agents' strategy update computes an equilibrium (resp. socially optimal) allocation based on the updated belief. Our treatment is related to the paper by Borkar and Kumar (2003) – their work focuses on stochastic approximation with two time-scales for analyzing the dynamics of cost estimates and asynchronous strategy updates in communication networks.

Our results are useful for designing an adaptive tolling mechanism to induce a socially optimal outcome in congestion games with unknown latency functions. For classical congestion games (i.e., full knowledge of latency functions), it is well-known that the negative externalities due to agents' selfish actions can be internalized by a tolling scheme based on marginal cost pricing Pigou (2017), Dial (1999), Roughgarden and Tardos (2002), and Ozdaglar and Srikant (2007). However, few have studied toll assignment under limited information about latency functions (Poveda et al. (2017), and Farokhi and Johansson (2015)). In Section 4, we present an adaptive pricing mechanism that computes belief-based toll assignments to ensure that the convergent load corresponds to a socially optimal outcome under complete information of latency functions.

2. Convergence Result and Fixed Point Properties

Our main result in this section is that the sequence of states $(\theta^k, x^k)_{k=1}^{\infty}$ converges to a fixed point $(\bar{\theta}, \bar{x})$ with probability 1. We also provide an asymptotic rate of convergence, and discuss some fixed point properties. To begin with, we introduce two basic definitions: fixed point of the learning dynamics (1), and set of distinguishable parameters based on an observed vector.

Definition 1 (Fixed point) *State $(\bar{\theta}, \bar{x})$ is a fixed point of (1) if*

$$\bar{\theta}(s) = \frac{\bar{\theta}(s)\widehat{\phi}^s(\hat{y}|\bar{x})}{\sum_{s' \in S} \bar{\theta}(s')\widehat{\phi}^{s'}(\hat{y}|\bar{x})}, \quad \forall s \in S, \quad \forall \hat{y} = (\hat{y}_e)_{e \in \{E|\bar{x}_e > 0\}}, \tag{2a}$$

$$\bar{x} = (1-a)\bar{x} + ag\left(\bar{\theta}\right), \quad \forall a \in (0,1]. \tag{2b}$$

That is, at a fixed point $(\bar{\theta}, \bar{x})$, the belief $\bar{\theta}$ is invariant to the Bayesian update (1a) for any randomly realized cost vector $\hat{y}$ on the resources utilized by $\bar{x}$ satisfying (2b). Note that if the learning dynamics starts at a fixed point, i.e. $(\theta^1, x^1) = (\bar{\theta}, \bar{x})$, then $(\theta^k, x^k) \equiv (\bar{\theta}, \bar{x})$ for any $k > 1$ w.p.1.

Definition 2 (Distinguishable parameters based on observed cost vector) *For a load vector $x \in X$, parameter s is distinguishable from the true parameter s^* based on the observed cost vector $\hat{y}$ if the Kullback–Leibler (KL) divergence between the distributions of $\hat{y}$ given s and s^* is positive, i.e.*

$$D_{KL}\left(\widehat{\phi}^{s^*}(\hat{y}|x)||\widehat{\phi}^{s}(\hat{y}|x)\right) = \int_{\hat{y}} \widehat{\phi}^{s^*}(\hat{y}|x) \log\left(\frac{\widehat{\phi}^{s^*}(\hat{y}|x)}{\widehat{\phi}^{s}(\hat{y}|x)}\right) d\hat{y} > 0.$$

The set of distinguishable parameters based on $\hat{y}$ is $\widehat{S}^{\dagger}(x) \triangleq \{S | D_{KL}\left(\widehat{\phi}^{s^*}(\hat{y}|x)||\widehat{\phi}^{s}(\hat{y}|x)\right) > 0\}$. It is well-known that the KL-divergence between any two distributions is non-negative, and is equal to zero if and only if the two distributions are identical (e.g., see Chapter 2 in Cover and Thomas (2012)). Therefore, if $s \in \widehat{S}^{\dagger}(x)$, then $\Pr(\widehat{\phi}^{s^*}(\hat{y}|x) \neq \widehat{\phi}^{s}(\hat{y}|x)) > 0$. Hence, an observed cost vector $\hat{y}$ based on load vector x can be used to distinguish $s \in \widehat{S}^{\dagger}(x)$ and s^*.

It is important to note that the set of distinguishable parameters in Definition 2 depends on the load vector x. A parameter $s \notin \widehat{S}^{\dagger}(x)$ may be distinguishable by another load vector that utilizes a different set of resources or utilizes resources with a different load level in comparison to x.

The following proposition characterizes the properties of fixed points:

Proposition 3 *Any state $(\bar{\theta}, \bar{x})$ such that $\bar{\theta}(s^*) > 0$ is a fixed point of (1) if and only if it satisfies:*

$$\bar{\theta}(s) = 0, \quad \forall s \in \widehat{S}^{\dagger}(\bar{x}), \tag{3a}$$

$$\bar{x} = g(\bar{\theta}). \tag{3b}$$

Hence, any fixed point $(\bar{\theta}, \bar{x})$ with positive belief on the true parameter s^* must assign zero probability to all the distinguishable parameters $s \in \widehat{S}^{\dagger}(\bar{x})$. Recall that the strategy update (1b) is a linear combination of x and $g(\theta)$, and the weight on $g(\theta)$ is positive; thus, a fixed point load vector $\bar{x}$ must be equal to $g(\bar{\theta})$. For such a fixed point, we must have that if $\bar{\theta}(s) > 0$ for a parameter $s \in S$, then $D_{KL}\left(\widehat{\phi}^{s^*}(\hat{y}|\bar{x})||\widehat{\phi}^{s}(\hat{y}|\bar{x})\right) = 0$; equivalently, $\widehat{\phi}^{s^*}(\hat{y}|\bar{x}) = \widehat{\phi}^{s}(\hat{y}|\bar{x})$ for any $\hat{y}$. Therefore, we can estimate the distribution of the observed cost vector $\hat{y}$ at fixed point $(\bar{\theta}, \bar{x})$:

$$\widehat{\mu}(\hat{y}|\bar{\theta}, \bar{x}) \triangleq \sum_{s \in S} \bar{\theta}(s)\widehat{\phi}^{s}(\hat{y}|\bar{x}) \overset{(3a)}{=} \sum_{s \in S \setminus \widehat{S}^{\dagger}(\bar{x})} \bar{\theta}(s)\widehat{\phi}^{s}(\hat{y}|\bar{x}) = \sum_{s \in S \setminus \widehat{S}^{\dagger}(\bar{x})} \bar{\theta}(s)\widehat{\phi}^{s^*}(\hat{y}|\bar{x}) = \widehat{\phi}^{s^*}(\hat{y}|\bar{x}). \tag{4}$$

In other words, using Proposition 3 we obtain that fixed point belief $\bar{\theta}$ must provide an accurate estimation of the observed cost distribution when the load vector is $\bar{x}$.

We are now ready to present the convergence theorem.

Theorem 4 *For any initial condition (θ^1, x^1), the sequence of states $(\theta^k, x^k)_{k=1}^{\infty}$ generated by the learning dynamics (1) converges to a fixed point $(\bar{\theta}, \bar{x}) \in \Delta(S) \times X$ with probability 1. Furthermore, for any $s \in \widehat{S}^{\dagger}(\bar{x})$, $\theta^k(s)$ converges to 0 exponentially fast:*

$$\lim_{k \to \infty} \frac{1}{k} \log(\theta^k(s)) = -D_{KL}(\widehat{\phi}^{s^*}(\hat{y}|\bar{x})||\widehat{\phi}^{s}(\hat{y}|\bar{x})). \quad w.p.1 \tag{5}$$

The proof of this result involves three steps: First, we use martingale convergence theorem to show that both sequences $\left(\frac{\theta^k(s)}{\theta^k(s^*)}\right)_{k=1}^{\infty}$ (this ratio is well-defined due to **(A2)**) and $\left(\theta^k(s^*)\right)_{k=1}^{\infty}$ converge with probability 1. Hence, the sequence of beliefs $\left(\theta^k\right)_{k=1}^{\infty}$ also converges to a belief $\bar{\theta}$ with probability 1. Second, by iteratively applying (1b), we can write x^k as a weighted summation of the initial

load vector x^1 and the sequence of $\left(g(\theta^j)\right)_{j=2}^{k+1}$. By utilizing the convergence of θ^k and **(A3)–(A4)**, we show that $\left(x^k\right)_{k=1}^{\infty}$ also converges to a fixed point load vector $\bar{x}$ with probability 1. Third, based on the convergence of both θ^k and x^k and **(A1)**, we argue that the log-likelihood ratio $\log\left(\frac{\theta^k(s)}{\theta^k(s^*)}\right)$ must converge to $-\infty$ for any distinguishable parameter $s \in \widehat{S}^\dagger(\bar{x})$. Therefore, the belief of any $s \in \widehat{S}^\dagger(\bar{x})$ converges to zero, with an exponential rate given by the (non-zero) KL-divergence between the distributions of observed cost vector under parameters s and s^*.

To summarize, the learning dynamics (1) converges to a fixed point with probability 1 (Theorem 4), and the belief distribution eventually forms an accurate estimation of the observed cost distribution with the fixed point load vector $\bar{x}$ (Proposition 3). However, it is important to note that the estimation of cost distribution may not be accurate on the resources that are not utilized by $\bar{x}$ (i.e. $e \in \{E|\bar{x}_e = 0\}$). Additionally, the estimation of cost distribution may not be accurate for a different load vector; i.e. when $x \neq \bar{x}$, the distribution $\hat{\mu}(\hat{y}|\bar{\theta}, x)$ may be different from $\phi^{s^*}(\hat{y}|x)$.

Next, we study how the initial load vector x^1 affects the properties of the fixed point. In particular, x^1 influences the fixed point belief $\bar{\theta}$ because it affects the costs of which resources are observed and incorporated in the belief update (1a). Thus, x^1 also affects the fixed point load vector $\bar{x}$ that must satisfy (2b). Proposition 7 clarifies how x^1 affects $(\bar{\theta}, \bar{x})$. As a preparation, we define the set of distinguishable parameters based on full cost vector y and introduce a lemma.

Definition 5 (Distinguishable parameters based on full cost vector) *For any $x \in X$, the set of distinguishable parameters from s^* based on y is $S^\dagger(x) \triangleq \{S|D_{KL}\left(\phi^{s^*}(y|x)||\phi^s(y|x)\right) > 0\}$, where $D_{KL}\left(\phi^{s^*}(y|x)||\phi^s(y|x)\right) = \int_y \phi^{s^*}(y|x)\log\left(\frac{\phi^{s^*}(y|x)}{\phi^s(y|x)}\right)dy$.*

Lemma 6 *$\forall x \in X$, $D_{KL}\left(\widehat{\phi}^{s^*}(\hat{y}|x)||\widehat{\phi}^s(\hat{y}|x)\right) \leq D_{KL}(\phi^{s^*}(y|x)||\phi^s(y|x))$ and $\widehat{S}^\dagger(x) \subseteq S^\dagger(x)$.*

Hence, any parameter s that is distinguishable from s^* based on $\hat{y}$ is also distinguishable based on y. The following proposition is a refinement of Theorem 4 for the case when all resources are utilized by the initial load vector, and the step size is strictly smaller than 1 for all steps.

Proposition 7 *If $x_e^1 > 0$ for all $e \in E$ and $a^k \in (0,1)$ for all k, then $(\theta^k, x^k)_{k=1}^{\infty}$ converges to a fixed point $(\bar{\theta}, \bar{x}) \in \Delta(S) \times X$ such that $\bar{\theta}(s) = 0$ for all $s \in S^\dagger(\bar{x})$ and $\bar{x} = g(\bar{\theta})$ with probability 1. Moreover, for any $s \in S^\dagger(\bar{x})$, $\theta^k(s)$ converges to 0 exponentially fast: $\lim_{k\to\infty}\frac{1}{k}\log(\theta^k(s)) = -D_{KL}(\phi^{s^*}(y|\bar{x})||\phi^s(y|\bar{x}))$.*

Under the conditions of Proposition 7, we can conclude that the belief $\bar{\theta}$ accurately estimates the distribution of costs on *all* resources, including the ones not utilized by $\bar{x}$. Since all the resources are utilized by x^1 and $a^k < 1$, the strategy update (1b) ensures that $x_e^k > 0$ for all $e \in E$ and all k. This is true even when fixed point load vector is such that there exists some $e \in E$ for which $\bar{x}_e = 0$ (i.e., $\lim_{k\to\infty} x_e^k = 0$), since such a resource is still utilized repeatedly in the learning dynamics. From **(A1)**, we obtain that any s that is distinguishable based on full cost vector given $\bar{x}$ is excluded from $\bar{\theta}$. Hence, the estimation of full cost distribution is accurate.

From Propositions 3 and 7, we obtain that when the learning dynamics starts with an initial load vector that utilizes all resources, the set of convergent states is a subset of all possible fixed points that are attainable from an arbitrary initial condition. In general, if a belief $\bar{\theta}$ forms an accurate estimation on resources that are utilized with $\bar{x}$ but not on the remaining ones, then $S^\dagger(\bar{x}) \setminus \widehat{S}^\dagger(\bar{x})$

must be a non-empty set, and there exists a $s \in S^\dagger(\bar{x}) \setminus \widehat{S}^\dagger(\bar{x})$ such that $\bar{\theta}(s) > 0$. Indeed, such a belief $\bar{\theta}$ would be a fixed point belief (i.e., it satisfies (3a) in Proposition 3), but cannot be a convergent belief of the dynamics that starts with all resources being utilized (Proposition 7).

Finally, from Theorem 4, Lemma 6, and Proposition 7, we conclude that when all resources are utilized initially and the step size is less than 1, the belief of the learning dynamics converges with a higher asymptotic rate, because the information on cost of resources that may not be utilized otherwise is included in the belief update in all steps.

3. Learning in congestion games with unknown cost parameter

In this section, we instantiate the general formulation of learning dynamics in Sec. 1 to a traffic routing (congestion) game. Specifically, E is a set of congestible resources, which form a network with multiple origin-destination (o-d) pairs belonging to the set I. Each o-d pair $i \in I$ is connected by a set of routes (i.e., sequence of resources) R_i. We denote $R = \cup_{i \in I} R_i$ as the set of all routes in the network. The cost of delay on each resource $e \in E$ is random and denoted by y_e. Importantly, the probability distribution of full cost vector $y = (y_e)_{e \in E}$ is governed by an unknown parameter $s \in S$.

A set of non-atomic agents make routing decisions on the network. The demand of agents routing between o-d pair $i \in I$ is $D_i \geq 0$, and the total demand is $D = \sum_{i \in I} D_i$. Let $f = (f_r)_{r \in R} \in F$ denote a routing strategy, where f_r is the traffic demand on route $r \in R$. A strategy f is feasible if $\sum_{r \in R_i} f_r = D_i$ for all $i \in I$, and $f_r \geq 0$ for all $r \in R$. For any resource $e \in E$, the load x_e is the sum of traffic flows on the routes passing through it, i.e. $x_e = \sum_{r \ni e} f_e$. The set of feasible load vectors (i.e. load vectors that can be induced by a feasible f) is convex, and denoted by X. For a given $x \in X$ and $s \in S$, the probability density function of cost vector y is $\phi^s(y|x)$.

For any $e \in E$, $s \in S$, and $x \in X$, we call the expected value of the realized cost y_e based on the probability density function $\phi^s(y|x)$ as the average cost ("latency") of the resource e under load x. A standard assumption in congestion games is that the latency function $\ell_e^s(\cdot) : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is an *increasing* function of load x_e for any $e \in E$ and $s \in S$. In addition, we assume that:

(A5) *For every $s \in S$ and $e \in E$, $\ell_e^s(x_e)$ is a strictly increasing C^2 function in x_e, and $(x_e \cdot \ell_e^s(x_e))$ is strictly convex in x_e. Additionally, $\exists \alpha, \beta > 0$ such that $\frac{d\ell_e^s(x_e)}{dx_e} \geq \alpha$, and $\frac{d^2(x_e \cdot \ell_e^s(x_e))}{dx_e^2} \geq \beta$ for all x_e , $e \in E$ and $s \in S$*

We now describe the learning dynamics (1) when non-atomic agents play the aforementioned traffic routing game in each step $k = 1, 2, \ldots$. Here the sequence of load vectors $\left(x^k\right)_{k=1}^{\infty}$ generated by strategy update (1b) capture the impact of agents' routing strategies on network congestion (i.e., level of utilization of various resources), and hence influence the realised costs in each step. The load vector x^k in each stage k is induced by the agents' routing strategy f^k as follows: for any function $g : \Delta S \to X$, one can find another function $f : \Delta(S) \to F$ that captures the agents' preferred routing strategy based on the belief θ; i.e., $g(\theta)$ gives the load vector induced by routing strategy $f(\theta)$. Thus, the dynamics of x^k in (1b) is induced by agents updating the routing strategy f^k following the dynamics $f^{k+1} = (1 - a^k) f^k + a^k f\left(\theta^{k+1}\right)$ for all k. Since the cost of each resource depends on the its aggregate load, we focus on the dynamics of x^k instead of f^k.

Furthermore, in congestion games, it is natural to focus on learning dynamics when the function $g(\theta)$ computes a Wardrop equilibrium or a socially optimal load vector, based on the current belief of cost parameter θ. A Wardrop equilibrium corresponds to the situation when agents selfishly prefer to minimize their expected individual cost of routing based on belief θ. On the other hand, a socially optimal load vector minimizes the expected social cost of all agents based on belief θ. Using known results on congestion games (Sandholm (2001)) and **(A5)**, we have the follows:

Lemma 8 *For any $\theta \in \Delta(S)$ and $x \in X$, let $\Phi(x|\theta) \triangleq \sum_{s\in S}\sum_{e\in E}\theta(s)\int_0^{x_e}\ell_e^s(z)dz$ and $\mathbb{E}[C(x)|\theta] \triangleq \sum_{s\in S}\sum_{e\in E}\theta(s)x_e\ell_e^s(x_e)$. Then, we have: **(i)** $g^{we}(\theta) = \arg\min_{x\in X}\Phi(x|\theta)$ is the unique equilibrium load vector. **(ii)** $g^{opt}(\theta) = \arg\min_{x\in X}\mathbb{E}[C(x)|\theta]$ is the unique socially optimal load vector. **(iii)** $g^{we}(\theta)$ and $g^{opt}(\theta)$ are continuous functions of θ.*

Note that both $\Phi(x|\theta)$ and $\mathbb{E}[C(x)|\theta]$ are convex functions of load vector x; hence, $g^{we}(\theta)$ and $g^{opt}(\theta)$ can be solved for any θ using known convex optimization algorithms.

Under assumptions **(A1) – (A3)** and **(A5)**, our results in Sec. 2 hold for learning dynamics (1) in the setting of congestion games. Specifically, with $g(\theta) = g^{we}(\theta)$ (resp. $g(\theta) = g^{opt}(\theta)$), the load vector eventually converges to $\bar{x}$, which is a Wardrop equilibrium (resp. socially optimal) load vector based on the convergent belief $\bar{\theta}$. This fixed point belief accurately estimates the cost distributions on resources that are utilized under $\bar{x}$ (Thm. 4). Additionally, when $x_e^1 > 0$ for all $e \in E$ and $a^k < 1$ for all k, the estimation of cost distribution on all resources is accurate (Prop. 7).

However, even under the conditions of Prop. 7, $\bar{\theta}$ may not accurately estimate the cost distribution when the underlying load vector is different from $\bar{x}$. The question then arises as to whether the fixed point condition $\bar{x} = g^{we}(\bar{\theta})$ (resp. $\bar{x} = g^{opt}(\bar{\theta})$) is equivalent to learning the Wardrop equilibrium (resp. socially optimal) load vector with complete information of the true cost parameter s^*, denoted x^{we*} (resp. x^{opt*}). Our next proposition addresses this question. We first introduce the following assumption on the latency functions, which is needed for the case of $g(\theta) = g^{opt}(\theta)$:

***(A6)** For all $s \in S$ and any $x > 0$, if $s \notin S^\dagger(x)$, then $\frac{d\ell_e^s(x_e)}{dx_e} = \frac{d\ell_e^{s^*}(x_e)}{dx_e}$ for all resources $e \in E$.*

If $s \notin S^\dagger(x)$ but $\frac{d\ell_e^s(x_e)}{dx_e} \neq \frac{d\ell_e^{s^*}(x_e)}{dx_e}$ on some resource $e \in E$, then perturbing x_e locally distinguishes s from s^*. Essentially, **(A6)** is weaker than the assumption that any s which is not distinguishable from s^* with load vector x is also not distinguishable in a small neighborhood of x.

Proposition 9 *Assume that **(A1) – (A3)** and **(A5)** hold. For learning dynamics (1) with $x_e^1 > 0$ for all $e \in E$ and $a^k \in (0,1)$ for all k, we have $\bar{x} \equiv x^{we*}$ for $g(\cdot) = g^{we}(\cdot)$. Additionally, we have $\bar{x} \equiv x^{we*}$ for $g(\cdot) = g^{opt}(\cdot)$ under **(A6)**.*

When $g(\cdot) = g^{we}(\cdot)$, the fixed point $\bar{\theta}$ accurately estimates the distribution of costs on all resources under load vector $\bar{x}$ (Prop. 7); thus the estimated value of $\ell_e^{s^*}(\bar{x}_e)$ must also be accurate on all resources. Then, we show that $\bar{x}$ satisfies the set of variational inequalities (Dafermos (1980)) with respect to the true cost parameter s^*, and hence must be the unique equilibrium load vector corresponding to the game with complete information of s^*.

On the other hand, when $g(\cdot) = g^{opt}(\cdot)$, then fixed point load vector $\bar{x} = g^{opt}(\bar{\theta})$ (the socially optimal load vector corresponding to $\bar{\theta}$). To obtain this conclusion, we show that $\bar{x}$ is the equilibrium load vector of a modified congestion game, where the latency function for each resource $e \in E$ and parameter $s \in S$ is $\ell_e^s(x_e) + x_e\left(d\ell_e^s(x_e)\right)/\left(dx_e\right)$. Then, from assumption **(A6)** and Prop. 7, we know that $\bar{\theta}$ accurately estimates the modified latency functions on all resources, and $\bar{x}$ is the equilibrium load vector of the modified congestion game, which is equivalent to x^{opt*}.

In fact, when the conditions in Prop. 9 are not satisfied, the fixed point load vector for learning dynamics (1) with $g(\cdot) = g^{we}(\cdot)$ (resp. $g(\cdot) = g^{opt}(\cdot)$) may not be equivalent to x^{we*} (resp. x^{opt*}), resulting in a higher social cost defined as $C^{s^*}(x) = \sum_{e\in E} x_e\ell_e^{s^*}(x_e)$ for $x \in X$. Specifically, our previous work Wu and Amin (2019) shows that when $g(\cdot) = g^{we}(\cdot)$ and the resource set E forms a series-parallel network (i.e. it does not have an embedded wheatstone network; see Milchtaich (2006))), we have $C^{s^*}(\bar{x}) \geq C^{s^*}(x^{we*})$ for any $\bar{x}$. Furthermore, since x^{opt*} minimizes $C^{s^*}(x)$, we directly have $C^{s^*}(\bar{x}) \geq C^{s^*}(x^{opt*})$ for all $\bar{x}$ in any network.

4. Adaptive learning of optimal toll assignment

In this section, we focus on adaptive learning of a socially optimal toll assignment by modifying the dynamics (1) to include toll assignment by a central authority based on the belief of unknown cost parameter. The augmented state in step k is $\left(\theta^k, \tau^k, x^k\right)$, where $\tau^k = \left(\tau_e^k\right)_{e\in E} \in \mathbb{R}_{\geq 0}^{|E|}$ is the vector of toll prices in step k. In each step k, the central authority updates the belief θ^k based on x^k and $\hat{y}^k$, uses the updated belief θ^{k+1} to revise τ^k, and announces τ^{k+1} to all agents; the agents update x^k induced by their routing strategy based on θ^{k+1} and τ^{k+1}.

To gain intuition about the modified dynamics, one can analyze how to assign toll for a given belief θ so that the induced equilibrium load is equivalent to the socially optimal load $g^{opt}(\theta)$. Let us assume that the toll τ_e on each resource has been converted from monetary price to the equivalent cost of delay. Then, for a load vector x, the cost experienced by agents utilizing resource e is $y_e + \tau_e$ (where y_e is the realized cost of delay) and the modified latency function is $\tilde{\ell}_e^s(x_e) = \ell_e^s(x_e) + \tau_e$. The equilibrium load vector can be computed as the minimizer of the potential function associated with the modified latency functions:

$$\tilde{g}^{we}(\theta, \tau) = \arg\min_{x\in X} \tilde{\Phi}(x|\theta) \triangleq \sum_{s\in S}\sum_{e\in E} \theta(s) \int_0^{x_e} \left(\ell_e^s(z) + \tau_e\right) dz. \tag{6}$$

If τ_e is set to $\sum_{s\in S} \theta(s) z \cdot \left(d\ell_e^s(z)\right) / \left(dz\right)$, which is the marginal cost of utilizing e, then the potential function $\tilde{\Phi}(x|\theta)$ is identical to the expected social cost $\mathbb{E}[C(x)|\theta] = \sum_{s\in S}\sum_{e\in E} \theta(s) x_e \ell_e^s(x_e)$. Then, $\tilde{g}^{we}(\theta, \tau) = \arg\min_{x\in X} \tilde{\Phi}(x|\theta) = \arg\min_{x\in X} \mathbb{E}[C(x)|\theta] = g^{opt}(\theta)$, which is the socially optimal load vector based on θ. The socially optimal toll that induces $g^{opt}(\theta)$ is $h(\theta) \triangleq \left(h_e(\theta)\right)_{e\in E}$:

$$h_e(\theta) = \sum_{s\in S} \theta(s) x_e \frac{d\ell_e^s(x_e)}{dx_e}\bigg|_{x=g^{opt}(\theta)}, \quad \forall e \in E. \tag{7}$$

Formally, the modified dynamics evolves as follows: In each step k, the belief θ^k is updated based on x^k and $\hat{y}^k$ according (1a); the toll vector τ^k and load vector x^k are updated as follows:

$$\tau^{k+1} = (1 - b^k)\tau^k + b^k h\left(\theta^{k+1}\right), \tag{8a}$$

$$x^{k+1} = (1 - a^k)x^k + a^k \tilde{g}^{we}\left(\theta^{k+1}, \tau^{k+1}\right), \tag{8b}$$

where $h(\cdot)$ and $\tilde{g}^{we}(\cdot)$ are given by (7) and (6), respectively. Thus, based on belief θ^k, (8a) linearly combines τ^k and the socially optimal toll vector with step size b^k. Based on θ^k and τ^k, (8b) linearly combines x^k and the equilibrium load vector with step size a^k. We modify **(A3)** as follows:

(A3′). *$a^k \in (0,1), b^k \in (0,1]$ for all k and $\prod_{k=1}^{\infty}(1-a^k) = \prod_{k=1}^{\infty}(1-b^k) = 0$.*

Finally, follows from Theorem 4, Lemma 8, and Proposition 9, we show that the load vector converges to socially optimal load vector x^{opt*} under complete information of cost parameter s^*, and the toll vector converges to the optimal toll vector $\tau^{opt*} = \left(x_e \frac{d\ell_e^{s^*}(x_e)}{dx_e}\Big|_{x_e = x_e^{opt*}}\right)_{e\in E}$.

Proposition 10 *Assume that **(A1)–(A2)**, **(A3′)**, and **(A5)–(A6)** hold. Then the modified dynamics (1a), (8a)-(8b) with $x_e^1 > 0$ for all $e \in E$ converges to a fixed point $\left(\bar{\theta}, \bar{\tau}, \bar{x}\right)$ with probability 1, where $\bar{\theta}(s) = 0$, for all $s \in S^\dagger(x^{opt*})$, $\bar{\tau} = \tau^{opt*}$, and $\bar{x} = x^{opt*}$.*

Thus, when the dynamics (1a), (8a)-(8b) starts with an initial load that utilizes all resources and the step size in each strategy update (8b) is less than 1, then at fixed point the belief provides an accurate estimation of costs on all resources and the toll is assigned to the socially optimal toll under full information of cost parameter, leading to a socially optimal load allocation.

References

Daron Acemoglu, Munther A Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

Vivek S Borkar and Panganamala Ramana Kumar. Dynamic Cesaro-Wardrop equilibration in networks. *IEEE Transactions on Automatic Control*, 48(3):382–396, 2003.

Vivek S Borkar and Sean P Meyn. The ODE method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469, 2000.

Roberto Cominetti, Emerson Melo, and Sylvain Sorin. A payoff-based learning procedure and its application to traffic games. *Games and Economic Behavior*, 70(1):71–83, 2010.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

Stella Dafermos. Traffic equilibrium and variational inequalities. *Transportation science*, 14(1): 42–54, 1980.

Robert B Dial. Network-optimized road pricing: Part I: A parable and a model. *Operations Research*, 47(1):54–64, 1999.

Farhad Farokhi and Karl H Johansson. A piecewise-constant congestion taxing policy for repeated routing games. *Transportation Research Part B: Methodological*, 78:123–143, 2015.

Drew Fudenberg and David M Kreps. Learning in extensive-form games I. self-confirming equilibria. *Games and Economic Behavior*, 8(1):20–55, 1995.

Douglas Gale and Shachar Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346, 2003.

Ali Jadbabaie, Pooya Molavi, and Alireza Tahbaz-Salehi. Information heterogeneity and the speed of learning in social networks. *Columbia Business School Research Paper*, (13-28), 2013.

Walid Krichene, Benjamin Drighes, and Alexandre Bayen. On the convergence of no-regret learning in selfish routing. In *International Conference on Machine Learning*, pages 163–171, 2014.

Igal Milchtaich. Network topology and the efficiency of equilibrium. *Games and Economic Behavior*, 57(2):321–346, 2006.

Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1): 124–143, 1996.

Asuman Ozdaglar and Rayadurgam Srikant. Incentives and pricing in communication networks. *Algorithmic Game Theory*, 647:571–591, 2007.

Arthur Pigou. *The economics of welfare*. Routledge, 2017.

Jorge I Poveda, Philip N Brown, Jason R Marden, and Andrew R Teel. A class of distributed adaptive pricing mechanisms for societal systems with limited information. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1490–1495. IEEE, 2017.

Benjamin Recht. A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:253–279, 2019.

Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2): 236–259, 2002.

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

Jeff S Shamma and Gürdal Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, 50(3):312–327, 2005.

John N Tsitsiklis. Asynchronous stochastic approximation and Q-learning. *Machine learning*, 16 (3):185–202, 1994.

Manxi Wu and Saurabh Amin. Learning an unknown network state in routing games. *arXiv preprint arXiv:1905.04433*, 2019.