Selfish Behavior and Resource Competition in Multi-Agent Systems

Costas Courcoubetis
Chinese University of Hong Kong
Shenzhen, China
costas@cuhk.edu.cn

Antonis Dimakis
Athens University of Economics and Business
Athens, Greece
dimakis@aueb.gr

ABSTRACT

We study the convergence and equilibrium behavior of a large number of selfish agents who interact by queuing for sequentially acquired consumable resources. Examples of such systems include ridehailing and crowdsourcing platforms, systems with energy-like resources such as charging stations, and communication systems. Despite the generality of the agents' Markov decision process structures, this type of interaction permits a tractable characterization of equilibria. In particular, we leverage the property that these equilibria can be formulated as optimal solutions to an extended Eisenberg-Gale program, where time serves as an analog for money.

Using this formulation, we (i) approximate equilibria via binary search, (ii) demonstrate Lyapunov stability for a broad class of learning dynamics, and (iii) establish global asymptotic stability of equilibria under replicator dynamics. Additionally, we prove Lyapunov stability for the coupled dynamics of queues and agents' replicator dynamics. When agents receive proportionally fair payoffs, they converge to an optimal set of actions, effectively behaving as if centrally coordinated.

KEYWORDS

Selfish agents; Resource competition; Proportional fairness; Learning dynamics

ACM Reference Format:
Costas Courcoubetis and Antonis Dimakis. 2025. Selfish Behavior and Resource Competition in Multi-Agent Systems. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.

1 INTRODUCTION

In this paper, we examine the behavior of noncooperative multi-agent systems composed of a large number of agents, each maximizing its own benefit. Such systems include, but are not limited to, ride-hailing, crowdsourcing, and communication networks. A common feature of these systems is that agents compete for resources required for task execution, which can lead to congestion effects.

There are two main approaches to studying multi-agent resource competition.

In the literature on congestion games and selfish routing, researchers have long studied properties of Nash (or Wardrop, in the nonatomic case) equilibria, starting with [1, 23]. The work by Roughgarden and Tardos [20] addresses efficiency questions in terms of the price of anarchy. The convergence to equilibrium has been tackled by a number of authors, e.g., by [14] in the atomic/discrete time case, and [8] in the nonatomic/continuous time case, where simple learning dynamics (related to replicator dynamics) are shown to converge. In these models, the number of agents varies over time as they enter or exit the system, making them less suitable for applications where the same agents interact over extended periods. Imposing 'agent mass conservation' constraints cannot work in these cases, as the variational characterizations foundational to these results will fail to hold.

In economics, Fisher's market is one of the most fundamental models of resource competition [17], where the numbers of agents and resources are fixed. Market-based models (see Jain and Vazirani [11] for generalizations of Fisher's market) are also frequently applied in noneconomic contexts, where prices can be interpreted as waiting times, as in Kelly's work [12]. In these settings, convergence to a competitive equilibrium is typically analyzed using tâtonnement processes or primal-dual algorithms (e.g., see [10]). More recently, proportional response dynamics –a straightforward adjustment that does not require gradient information– has been shown to converge in discrete time for Fisher's markets [2, 24], with competitive prices instantly recalculated after each adjustment. In noneconomic settings, it is more appropriate to consider gradual price/waiting time updates as these are driven by the natural queueing dynamics (which are of the tâtonnement type). More importantly, convergence of proportional response has been analyzed under Fisher's market constraints, where resources are independent. However, in many real-world applications, agents' actions require multiple resources simultaneously, leading to interdependent constraints.

In this paper, we consider a nonatomic model of resource competition using a Markov Decision Process (MDP) formalism. We define equilibrium as a market-based equilibrium in a noneconomic context, where prices represent time. Unlike in traditional economic applications, our notion of equilibrium is not competitive, as even undesired actions may have positive execution times[1]. Nevertheless, we show that equilibria in this type of market correspond to optimal solutions of an extended Eisenberg-Gale program [7]. We leverage this connection to approximate equilibria through a straightforward algorithm that combines binary search with a network flow problem. We note that markets with interdependent constraints, such as the ones we consider in this work, have fundamentally different characteristics compared to Fisher's markets. For example, there can be non-determinacy for the waiting times that form in

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

[1] In a competitive market equilibrium, undesired resources have zero prices.

the market, which complicates the definition of equilibria and the rest of the analysis (potential function, learning dynamics, etc).

Using the set of deterministic unichain[2] policies, we study convergence of learning dynamics to equilibrium. Although the potential function we use is related to the work in [2, 5], we derive it independently not via Shmyrev's problem [22] as in those references, but through an Eisenberg-Gale type of program distinct from the one used for equilibria characterization. This new potential formulation enables us to establish Lyapunov stability for a broad range of dynamics and prove global asymptotic stability for replicator dynamics. Furthermore, we demonstrate Lyapunov stability when agents follow replicator dynamics and queues adjust according to a natural tâtonnement mechanism, an outcome derived directly from our potential formulation.

Our work can also be seen as an extension of fluid models for multi-class closed queueing networks [13, 15, 21], reframing them within a game-theoretic context. In these models, there is a closed network of queues where a finite population of customers of different types circulate continuously, each type corresponding to a probabilistic queue routing profile, and customers interacting through their waiting in the queues. Our model is a nonatomic version of the above system that associates rewards with completing service at different subsets of the queues, with customers free to switch types if they find it more profitable in the long run. More specifically, in our model, agents subscribe to 'policies', i.e., a type of randomized cyclic behaviour, and collect the corresponding long-run average reward equal to the average reward of the cycle divided by the average cycle time. In an equilibrium, all cycles with positive agent mass have same average rewards, and convergence to the equilibrium occurs by agents switching to cycles that generate higher average revenue.

The paper is organized as follows. Section 2 introduces the basic model and notation. In Section 3, we define the concept of equilibrium and present Theorem 1, which characterizes equilibria as optimal solutions to an extended Eisenberg-Gale program. Section 4 discusses the polynomial-time approximation of equilibrium. Section 5 addresses learning dynamics: Proposition 1 offers an alternative equilibrium characterization in policy space using a potential function, which underpins the main convergence results. These include Proposition 2 on Lyapunov stability for a range of 'sensible' dynamics, Theorem 2 on replicator dynamics convergence, and Proposition 3 on the Lyapunov stability of joint replicator and queueing (tâtonnement) dynamics. In Section 6, we consider ridehailing as an application of our model and demonstrate the behavior of the joint dynamics using a numerical example based on data from the NYC area [6]. A summary is given in Section 7. Proofs not included in the main sections are provided in the Appendix.

2 MODEL AND DEFINITIONS

2.1 Model of a Single Agent

A nonatomic set of agents of unit total mass compete for resources. The state of each agent evolves according to a continuous time semi-Markov decision process with finite state space S and finite set of actions, A. The actions are chosen at state transition instances so that the choice of $a \in A$ in $i \in S$ results in the following sequence:

(1) An immediate reward r_{ia} is awarded[3] to the acting agent.
(2) The next state of the acting agent is chosen independently of the past according to the transition probabilities p^a_{ij}, $i, j \in S, a \in A$.
(3) The transition to the next state occurs after a random time with mean τ_{ia} (when averaged also over the next state j). We refer to τ_{ia} as the *mean sojourn time* in state-action pair (i, a). Its precise form is given in the following subsection.
(4) Upon entry to j the agent decides the next action and the process continues as above.

A *(deterministic) policy* σ is a function $\sigma : S \to A$. Let X_n be the state visited at the n-th transition for $n = 0, 1, \ldots$[4], and let the action selected at that instant be $\sigma(X_n)$, for some policy σ. For any starting state i, the *average reward* is

$$V(i, \sigma, \tau) = \liminf_T \frac{1}{T} E\left(\sum_{n=1}^{N_T} r_{X_n \sigma(X_n)} \middle| X_0 = i \right),$$

where $\tau = (\tau_{ia}, i \in S, a \in A)$, and N_T is the number of transitions before time T.

We will assume the MDP is *weakly communicating* [19], i.e., the state-space can be partitioned into sets S_0, S_1 where the states in S_0 are transient under any policy, and every state in S_1 is accessible from any other in S_1 under some policy. This makes the optimal average case reward independent of the starting state.

A policy σ_* is *optimal* if

$$V(i, \sigma_*, \tau) = \sup_\sigma V(i, \sigma, \tau) \equiv V_*(\tau) \text{ for all } i. \tag{1}$$

We will restrict[5] ourselves to the set of deterministic policies Π possessing a single recurrent class, i.e., the set of *unichain policies*.

2.2 Resource competition

When choosing an action, an agent must obtain the resources required for its execution. Thus, we assume τ_{ia} is decomposed as

$$\tau_{ia} = t_{ia} + \sum_{l=1}^{L} \alpha_{l,ia} w_l,$$

where $t_{ia} > 0$ is a constant *action execution time* and the second term represents the *waiting time* to collect the resources required by action a in i. Here, l indexes the set of resources $\{1, \ldots, L\}$, $\alpha_{l,ia}$ denotes the units of resource l required by action a in state i, and w_l denotes the waiting time to obtain one unit of resource l. Additionally, let the constant supply rate of resource l be $b_l > 0$, for $l = 1, \ldots, L$. To emphasize the dependence of τ_{ia} on $w = (w_l, l = 1, \ldots, L)$, we write $\tau_{ia}(w) \equiv \tau_{ia}$.

The waiting times are only assumed to satisfy the 'fluid queue' condition: there is no waiting if the rate resource l is consumed is strictly below its supply rate, i.e.,

$$\sum_{i,a} \alpha_{l,ia} x_{ia} < b_l \implies w_l = 0, \tag{2}$$

[2]The policies for which the resulting Markov chain has a single recurrent class of states (plus possibly some transient states).

[3]Nonpositive rewards are allowed.

[4]X_0 is the initial state.

[5]This restriction does not yield inferior policies since for any optimal deterministic policy $\sigma \notin \Pi$, a $\sigma' \in \Pi$ with $V(i, \sigma', \tau) = V(i, \sigma, \tau)$ can be constructed by taking $\sigma' = \sigma$ on a recurrent class S_σ of σ, and making σ' move towards S_σ elsewhere.

where x_{ia} is the agent mass per unit time, or *rate*, entering state i and choosing action a.

3 EQUILIBRIUM

We are now ready to define our main equilibrium concept.

Definition 1 (Equilibrium). *A pair (x, w) of rates $x = (x_{ia}, i \in S, a \in A)$ and resource waiting times $w = (w_l, l = 1, \ldots, L)$ is an* equilibrium *if they satisfy the following:*

(1) Resource constraints:

$$\sum_{i,a} \alpha_{l,ia} x_{ia} \le b_l,\ l = 1, \ldots, L. \tag{3}$$

(2) Fluid queue condition: (2) *holds for every $l = 1, \ldots, L$.*

(3) Flow balance:

$$\sum_{a \in A} x_{ia} = \sum_{j \in S, a \in A} x_{ja} p^a_{ji},\ \ \textit{for each } i \in S. \tag{4}$$

(4) Conservation of mass:

$$\sum_{i \in S, a \in A} \tau_{ia}(w) x_{ia} = 1. \tag{5}$$

(5) Individual optimality: *Any $\sigma \in \Pi$ with $x_{i\sigma(i)} > 0$ for all i in its recurrent class is optimal, i.e., $V(i, \sigma, \tau(w)) = V_*(\tau(w))$ for all i.*

The first four conditions pertain to the behavior of the aggregate, while the fifth condition relates to the selfish behavior of individual agents: if a non-negligible subset of agents uses action a in state i, it is because it maximizes the average reward of an individual.

To establish Theorem 1, as a first step we obtain an equivalent description of an equilibrium which uses the linear program (LP) form of the average case dynamic program (e.g., see [19]):

Lemma 1. *(x, w) is an equilibrium if and only if it satisfies the resource constraints* (3)*, the fluid queue condition* (2)*, and x is an optimal solution of the LP:*

$$\max \sum_{i,a} r_{ia} x_{ia} \tag{6}$$

$$\textit{s.t. flow balance } (4), \textit{ and} \tag{7}$$

$$\textit{mass conservation } (5) \textit{ hold}, \tag{8}$$

$$\textit{over } x = (x_{ia}, i \in S, a \in A) \in \mathbb{R}^{S \times A}_+.$$

Proof. If (x, w) is an equilibrium, x is a feasible solution of the LP. We will show that it is an optimal solution.

First notice that in the case of a weakly communicating chain there exists $h : S \to A$ which satisfies the dynamic programming equation:

$$h(i) = \max_{a \in A} \left(r_{ia} - \tau_{ia} V_*(\tau) + \sum_{j \in S} p^a_{ij} h(j) \right), \tag{9}$$

for all $i \in S$. If $x_{ia} > 0$ then there exists an optimal policy σ with $\sigma(i) = \alpha$, so

$$h(i) = r_{ia} - \tau_{ia} V_*(\tau) + \sum_{j \in S} p^a_{ij} h(j). \tag{10}$$

Thus,

$$\sum_{i,a} r_{ia} x_{ia} = \sum_{i,a} \left[r_{ia} x_{ia} + x_{ia} \left(h(i) - r_{ia} + \tau_{ia} V_*(\tau) - \sum_{j \in S} p^a_{ij} h(j) \right) \right]$$

$$= V_*(\tau) \sum_{i,a} x_{ia} \tau_{ia} + \sum_i h(i) \left(\sum_a x_{ia} - \sum_{j,a} p^a_{ji} x_{ja} \right) = V_*(\tau),$$

by conditions 3 and 4 in Definition 1. Since $V_*(\tau)$ is the optimal value of (6), we conclude that x is an optimal solution.

For the converse, let h satisfy (9) and notice that

$$\sum_{i,a} x_{ia} \left(h(i) - r_{ia} + \tau_{ia}(w) V_*(\tau(w)) - \sum_j p^a_{ij} h(j) \right) =$$

$$\sum_i h(i) \left(\sum_a x_{ia} - \sum_{j,a} p^a_{ji} x_{ja} \right) = 0,$$

since $\sum_{i,a} r_{ia} x_{ia} = V_*(\tau) = \sum_{i,a} x_{ia} \tau_{ia} V(r, p, \tau)$ by the optimality of x. Thus, (10) holds for all i with $x_{ia} > 0$ for some $a \in A$, and so any policy σ as in the statement is optimal. □

This equivalent definition in Lemma 1 is analogous to the concept of equilibrium in Fisher's markets, where w is interpreted as a competitive price vector and (5) as a budget constraint. However, unlike in Fisher's markets, agents are not charged directly according to w; instead, they pay $\tau_{ia}(w)$, which includes an additional flat fee due to constant action execution times. Nevertheless, it is not surprising that the equilibria are characterized by a convex program that extends the Eisenberg-Gale program.

Theorem 1. *(x, w) is an equilibrium if and only if x is an optimal solution of:*

$$\max \quad \log \left(\sum_{i,a} r_{ia} x_{ia} \right) - \sum_{i,a} t_{ia} x_{ia} \tag{11}$$

$$\textit{s.t. } \sum_{i,a} \alpha_{l,ia} x_{ia} \le b_l,\ , l = 1, \ldots, L, \tag{12}$$

$$\sum_a x_{ia} = \sum_{j,a} x_{ja} p^a_{ji},\ i \in S, \tag{13}$$

$$\textit{over } x = (x_{ia}, i \in S, a \in A) \in \mathbb{R}^{S \times A}_+,$$

and $w = (w_l, l = 1, \ldots, L)$ are optimal Lagrange multipliers for the resource constraints (12)*.*

Proof of Theorem 1. If (x, w) is an equilibrium, then Lemma 1 implies that x maximizes $m \log \left(\sum_{i,a} r_{ia} x_{ia} \right)$ under the constraints (4), (5). Thus, there exist Lagrange multipliers $\nu \in \mathbb{R}, \xi \in \mathbb{R}^S$ for which

$$\frac{r_{ia}}{\sum_{j,\beta} r_{j\beta} x_{j\beta}} - \nu \tau_{ia}(w) - \xi_i + \sum_j p^a_{ij} \xi_j \le 0, \tag{14}$$

with the equality holding if $x_{ia} > 0$, so

$$\sum_{i,a} x_{ia} \left(\frac{r_{ia}}{\sum_{j,\beta} r_{j\beta} x_{j\beta}} - \nu \tau_{ia}(w) - \xi_i + \sum_j p^a_{ij} \xi_j \right) = 0.$$

Applying (7) and (8) to the last equality yields $\nu = 1$. Consequently, (14) holds with $\nu = 1$, and in combination with the feasibility conditions (3), (4) and the complementary slackness condition (2), this implies the optimality of x and the dual optimality of w, ξ.

For the converse, let x be an optimal solution, and w be an optimal set of Lagrange multipliers. Making use of (13) in the optimality condition (14) (with $\nu = 1$) implies (5) holds, and so x is feasible in (6). Since also (14) holds for $\nu = 1$, x satisfies the optimality conditions for (6). This and the fact that (x, w) satisfies (2), (3), implies (x, w) is an equilibrium, by Lemma 1. □

The assumption $t_{ia} > 0$ for all i, a implies an optimal solution of (11) exists, guaranteeing the existence of an equilibrium. Moreover, the strict concavity of the objective with respect to the average reward $\sum_{i,a} r_{ia}x_{ia}$ implies that its value in equilibrium is unique.

Corollary 1. *An equilibrium always exists, and the average reward is the same across all equilibria.*

4 EQUILIBRIUM COMPUTATION

In this section we give a polynomial time approximation algorithm for computing an equilibrium, based on binary search.

Observe that problem (11) is equivalent to

$$\begin{aligned} \max \quad & m \log\left(\sum_{i,a} r_{ia}x_{ia}\right) - m_o && (15)\\ \text{s.t.} \quad & \sum_{i,a} \alpha_{l,ia}x_{ia} \le b_l,\ l = 1,\dots,L, \\ & \sum_{a} x_{ia} = \sum_{j,a} x_{ja}p^a_{ji},\ i \in S, \\ & \sum_{i,a} t_{ia}x_{ia} \le m_o, && (16)\\ & \text{over } x_{ia} \ge 0, i \in S, a \in A, m_o \ge 0, \end{aligned}$$

where the variable m_o represents the mass of *active* agents: at the optimal solution, where (16) holds with equality, m_o equals the mass of agents executing some action, or equivalently, not waiting in a queue.

Optimizing over the x_{ia}'s first and then over m_o, allows to rewrite (15) as

$$\max_{m_o \ge 0} m \log F(m_o) - m_o, \tag{17}$$

where $F(m_o)$ is the optimal value of the LP:

$$\begin{aligned} \max \quad & \sum_{i,a} r_{ia}x_{ia} && (18)\\ \text{s.t.} \quad & \sum_{i,a} \alpha_{l,ia}x_{ia} \le b_l,\ l = 1,\dots,L, \\ & \sum_{a} x_{ia} = \sum_{j,a} x_{ja}p^a_{ji},\ i \in S, \\ & \sum_{i,a} t_{ia}x_{ia} \le m_o, && (19)\\ & \text{over } x_{ia} \ge 0, i \in S, a \in A. \end{aligned}$$

Since $F(m_o)$ is a concave function of m_o, the objective function in (17) is strictly concave so its maximum is achieved at the unique m_o for which $mF'(m_o) = F(m_o)$ holds, for some subgradient $F'(m_o)$ of $F(\cdot)$ at m_o. This m_o can be approximated arbitrarily well using binary search. This is the idea of Algorithm 1 which runs in polynomial time: an LP is solved in each iteration and binary search requires $\log(1/\epsilon)$ steps for approximation within ϵ.

In each iteration, a subgradient of $F'(m_o)$ is calculated, using any optimal Lagrange multiplier ν of the constraint (19). Because of the concavity of $F(\cdot)$, if the search continues in the left (right) interval then ν can be used as a lower (upper) bound of the range of subgradients $[\nu^-, \nu^+]$ for the next iteration. For this reason, the comparison needs to only consider a single subgradient ν, thus avoiding the computation of the entire subgradient range.

Algorithm 1 Equilibrium computation

Input: Active mass approximation tolerance $\epsilon > 0$.
Output: $x_{ia}, i \in S, a \in A$.

Let $a \leftarrow 0, b \leftarrow 1, \nu^+ = +\inf, \nu^- = 0$
while $b - a > \epsilon$ **do**
  $m_o \leftarrow \frac{a+b}{2}$
  $(F(m_o), x, \nu) \leftarrow$ (optimal value of (18), optimal solution, Lagrange multiplier of (19))
  if $\frac{F(m_o)}{m} > \nu$ **then**
    $a \leftarrow m_o$
    $\nu^+ \leftarrow \nu$
    $x^+ \leftarrow x$
  else if $\frac{F(m_o)}{m} < \nu$ **then**
    $b \leftarrow m_o$
    $\nu^- \leftarrow \nu$
    $x^- \leftarrow x$
  else
    break
  end if
end while
$\theta \leftarrow \frac{F(m_o)-\nu^-}{\nu^+-\nu^-}$
$x \leftarrow \theta x^+ + (1-\theta)x^-$
return $x = (x_{ia}, i \in S, a \in A)$.

5 LEARNING DYNAMICS

In this section we consider whether an agent population will evolve towards an equilibrium under simple learning dynamics.

To better understand learning dynamics and convergence, we use an alternative formulation of the agent decision problem. Instead of making separate decisions about actions in each state as the MDP formulation suggests, agents subscribe to policies from the set of unichain policies Π defined in Section 2.1. Each such policy fully specifies all the action choices of the agent at the different states of the MDP, and corresponds to an ergodic Markov chain over the state-actions, i.e., *a randomized cyclic behaviour* in the corresponding closed resource network that gets renewed each time some target state is visited. There are finitely many such 'policy cycles' available for agents to choose from, since Π is finite. Given the state of the system, i.e., the waiting times in the queues, each such policy offers a certain average reward per renewal cycle and requires a certain average cycle time. Agents prefer to switch to policies that offer a higher rate of average reward (cycle reward divided by cycle time). Eventually, equilibria will form where only policies with the highest possible (hence equal) reward rates attract a positive agent mass.

It is important to highlight that there is an important aspect of non-determinacy in this model that we must deal with. Policies that are not chosen at a given time by a positive mass of agents have not unique average cycle execution times, making many revenue

rates possible for these policies. This is because the waiting times that form in the resources can take many possible combinations of values. But, interestingly enough, for cycles with a positive mass of agents, the corresponding cycle times, and hence the revenue rates, are always uniquely determined.

This non-determinacy aspect for unused policies is taken into account in our definition of equilibrium (by introducing explicitly the waiting times at the queues in addition to the distribution of agents to policies), and in the analysis of the learning dynamics.

Before turning into this, we first characterize the rates and waiting delays for *any* selection of policies in Π by the agents, not just for the equilibrium, using a convex program ((23) below).

5.1 Policy-space formulation

For every $\sigma \in \Pi$, let m_σ be the mass of agents using policy σ. Define the *policy rate* x_σ as the expected rate of transitions of agents following σ, i.e.,

$$x_\sigma = \frac{m_\sigma}{\tau_\sigma(w)}, \sigma \in \Pi, \tag{20}$$

where $w = (w_\sigma, \sigma \in \Pi)$ is the vector of waiting delays, $\tau_\sigma(w) = \sum_{i,a} \pi^\sigma_{ia} \tau_{ia}(w)$ is the expected time between transitions for agents following policy σ, and $\pi^\sigma = (\pi^\sigma_{ia})$ is the invariant distribution of state-actions of the embedded chain of jumps (also under σ), i.e., the solution of $\sum_a \pi^\sigma_{ia} = \sum_{j,a} \pi^\sigma_{ja} p^a_{ji}$ for all i. Of the transitions counted in (20), those that leave i with action a having been selected come at a rate of $x_\sigma \pi^\sigma_{ia}$. Thus, the state-action rates can be written in terms of the policy rates:

$$x_{ia} = \sum_{\sigma \in \Pi} x_\sigma \pi^\sigma_{ia}, i \in S, a \in A. \tag{21}$$

(Conversely, for any selection of state-action rates which satisfy (4) there exist policy rates for which (21) holds, e.g., see Corollary 8.8.7 in [19].) Next, we show that the policy rates are defined uniquely for each probability distribution on Π given by $m = (m_\sigma)$.

Let $W(m)$ denote the set of $w = (w_l)$ for which the induced policy rates (x_σ) satisfy

$$\sum_{\sigma,i,a} \pi^\sigma_{ia} \alpha_{l,ia} x_\sigma \le b_l, \tag{22}$$

and $w_l = 0$ if the inequality is strict, for every $l = 1, \dots, L$.

Lemma 2. *For each $m = (m_\sigma)$, the policy rates are unique, i.e., x_σ in (20) attains the same value for all $w \in W(m)$. In particular, (x_σ) is the optimal solution of*

$$\max \sum_{\sigma: m_\sigma > 0} m_\sigma \log \frac{r_\sigma x_\sigma}{m_\sigma} - \sum_{\sigma,i,a} \pi^\sigma_{ia} t_{ia} x_\sigma \tag{23}$$

$$s.t. \sum_{\sigma,i,a} \pi^\sigma_{ia} \alpha_{l,ia} x_\sigma \le b_l, \quad l = 1, \dots, L, \tag{24}$$

$$over\ x_\sigma \ge 0, \sigma \in \Pi,$$

while $W(m)$ is the set of optimal dual variables corresponding to (24).

To highlight the dependence of the policy rates on m, we denote x_σ as $x_\sigma(m)$.

Equation (20) implies that $\tau_\sigma(w)$ is unique too if $m_\sigma > 0$. Since the average reward $u_\sigma(w)$ of σ (viz. the expected reward per unit time) can be expressed as the ratio $r_\sigma / \tau_\sigma(w)$ of the expected reward per transition, $r_\sigma = \sum_{i,a} \pi^\sigma_{ia} r_{ia}$, and the expected time between transitions, it is unique as well, if $m_\sigma > 0$.

Corollary 2. *If $m_\sigma > 0$ the values of $\tau_\sigma(w), u_\sigma(w)$ are the same across all $w \in W(m)$.*

Let $\phi(m)$ be the optimal value of (23), and its dual,

$$\min \sum_\sigma m_\sigma \Big(\log u_\sigma(w) - 1 \Big) + \sum_l b_l w_l \tag{25}$$

$$\text{over } w_l \ge 0,\ l = 1, \dots, L.$$

The distributions of agents that correspond to equilibria achieve the maximum value of $\phi(m)$ over all probability distributions m. To show this we will work with an alternative problem which is based on the dual of max $\phi(m)$:

$$\min \sum_l b_l w_l + v \tag{26}$$

$$\text{s.t. } \log u_\sigma(w) - 1 \le v,\ \sigma \in \Pi,$$

$$\text{over } w \ge 0, v.$$

Lemma 3. *m is a maximizer of $\phi(m)$ over the probability distributions on Π if and only if it is an optimal dual variable in (26).*

Proposition 1. *Let w, m be a primal and dual optimal solution, respectively, of (26), and $x = (x_{ia})$ be the corresponding state-action rates (through (21)). The pair (x, w) is an equilibrium.*

Conversely, let (x, w) be an equilibrium, and $x_\sigma, \sigma \in \Pi$ any set of corresponding policy rates. Let m be the probability distribution defined by $m_\sigma = x_\sigma \tau_\sigma(w)$ for every $\sigma \in \Pi$. Then w, m is a primal and dual optimal solution, respectively, of (26).

5.2 Stability and Convergence

While our results could also be formulated in discrete time, we use continuous-time dynamics to circumvent stability issues associated with step-size selection, as reported in [3] and [18].

We consider dynamics of the form $\dot m_\sigma(t) = G_\sigma(m(t)), \sigma \in \Pi, t \ge 0$, where the functions $G_\sigma : \mathbb{R}^\Pi_+ \to \mathbb{R}, \sigma \in \Pi$ satisfy the properties:

Properties. *(1) Lipschitz continuity,*

(2) Conservation of mass: $\sum_\sigma G_\sigma(m) = 0$.

(3) Forward invariance: $m_\sigma(t) = 0$ if and only if $m_\sigma(0) = 0$. (E.g., this holds if $m_\sigma = 0 \Rightarrow G_\sigma(m) = 0, \sup_{m_\sigma > 0} \frac{|G_\sigma(m)|}{m_\sigma} < \infty$ [16].)

(4) If $G_\sigma(m) > 0 > G_{\sigma'}(m) \Rightarrow u_\sigma(w) > u_{\sigma'}(w)$ for every $w \in W(m)$.

Lipschitz continuity is needed for the existence and uniqueness of trajectories; forward invariance implies that extinct strategies do not resurface and conversely, initially existent strategies do not become extinct in finite time. Property 4 states that if more agents play σ and less play σ' then the payoff of playing σ must be strictly greater than that of σ'. Most sensible policies have this property, e.g., best response, replicator, Brown-von Neumann-Nash, logit dynamics [9].

Note that policies that depend on the average reward $u_\sigma(w)$ are included in the above framework as their dynamics depend only on m. This is because $u_\sigma(w)$ takes the same value for all $w \in W(m)$ if $m_\sigma > 0$, by Corollary 2; if $m_\sigma = 0$ forward invariance implies $G_\sigma(m) = 0$ anyway.

For any subset of probability distributions M on Π, let $B_\delta(M)$ be 'ball of radius δ around M', i.e., the set of probability distributions m with $\min_{m' \in M} d(m, m') \le \delta$, where $d(\cdot, \cdot)$ is the Euclidean distance.

The set of distributions M is said to be *Lyapunov stable* if for every $\epsilon > 0$ there exists $\delta > 0$ such that $m(0) \in B_\delta(M)$ implies $m(t) \in B_\epsilon(M)$ for all $t \geq 0$; in other words, trajectories remain close to M when they start close to it.

Proposition 2. *The set M^o, consisting of the maximizers of $\phi(m)$, is Lyapunov stable under learning dynamics that satisfy Properties 1-4.*

Proof. Let m^o be any maximizer of $\phi(m)$, and for any $\epsilon > 0$ let V_ϵ be the set of probability distributions m with $\phi(m) \geq \phi(m^o) - \epsilon$.

We first show that V_ϵ is invariant, i.e., if $m(0) \in V_\epsilon$ then $m(t) \in V_\epsilon$ for all $t \geq 0$. Applying Danskin's envelope theorem [4] to (25) gives

$$\begin{aligned}\dot{\phi}(m(t)) &= \min_{w \in W(m(t))} \sum_\sigma G_\sigma(m(t)) \left(\log u_\sigma(w) - 1\right) \\ &= \sum_\sigma G_\sigma(m(t)) \left(\log u_\sigma(w) - 1\right),\end{aligned}$$

for every $w \in W(m(t))$, by Corollary 2. If $m(t)$ is not a stationary point for the dynamics then the sets $P = \{\sigma : G_\sigma(m(t)) > 0\}, N = \{\sigma : G_\sigma(m(t)) < 0\}$ are nonempty (by Property 2), and

$$\begin{aligned}\dot{\phi}(m(t)) &= \sum_{\sigma \in P} G_\sigma(m) \left(\log u_\sigma(w) - 1\right) + \sum_{\sigma \in N} G_\sigma(m) \left(\log u_\sigma(w) - 1\right) \\ &\geq \sum_{\sigma \in P} G_\sigma(m) \min_{\sigma' \in P} \left(\log u_{\sigma'}(w) - 1\right) \\ &\quad + \sum_{\sigma \in N} G_\sigma(m) \max_{\sigma' \in N} \left(\log u_{\sigma'}(w) - 1\right) \\ &> \sum_{\sigma \in P} G_\sigma(m) \max_{\sigma' \in N} \left(\log u_{\sigma'}(w) - 1\right) \\ &\quad + \sum_{\sigma \in N} G_\sigma(m) \max_{\sigma' \in N} \left(\log u_{\sigma'}(w) - 1\right) = 0,\end{aligned}$$

where Property 4 is used in the second inequality, and Property 2 in the final equality. If $m(t)$ is a stationary point then $\dot{\phi}(m(t)) = 0$, so $\dot{\phi}(m(t)) \geq 0$ in any case, i.e., $m(t)$ moves along ascending directions of ϕ. Thus, if $m(0) \in V_\epsilon$ then $\phi(m(t)) \geq \phi(m(0))$, implying $m(t) \in V_\epsilon$ for all $t \geq 0$, i.e., V_ϵ is invariant.

Because ϕ is continuous, choose $\delta > 0$ such that $V_\delta \subset B_{\frac{\epsilon}{2}}(M^o)$. Then $m(0) \in V_\delta$ implies $m(t) \in V_\delta \subset B_{\frac{\epsilon}{2}}(M^o) \subset B_\epsilon(M^o)$, i.e., the set M^o is Lyapunov stable. □

More can be shown for the replicator dynamics with fitness function $h_\sigma(u) = \log u_\sigma$, for every $u \in \mathbb{R}^\Pi_{++}$:

$$G_\sigma(m) = m_\sigma \left(h_\sigma(u_\sigma(w)) - \sum_{\sigma'} m_{\sigma'} h_{\sigma'}(u_\sigma(w)) \right), \ \sigma \in \Pi, m \in \mathbb{R}^\Pi_+. \tag{27}$$

In Theorem 2, we show convergence to an equilibrium for any initial distribution $m(0)$.

Theorem 2. *Assume the starting policy mix $m(0)$ contains an equilibrium, i.e., there exists probability distribution m^o supported in the support of $m(0)$ for which the corresponding state-action rates x paired with a $w^o \in W(m^o)$ is an equilibrium. Then, the trajectory $m(t), t \geq 0$ under replicator dynamics (27) converges to an equilibrium.*

Proof of Theorem 2. For any $t \geq 0$, let

$$D_{\mathrm{KL}}\left(m^o || m(t)\right) = -\sum_\sigma m^o_\sigma \log \frac{m_\sigma(t)}{m^o_\sigma},$$

denote the Kullback-Leibler divergence between $m(t), m^o$, which is well-defined since $\mathrm{supp}(m^o) \subseteq \mathrm{supp}(m(0)) = \mathrm{supp}(m(t))$. Then,

$$\begin{aligned}\dot{D}_{\mathrm{KL}}\left(m^o || m(t)\right) &= -\sum_\kappa m^o_\sigma \frac{\dot{m}_\sigma(t)}{m_\sigma(t)} \\ &= -\sum_\sigma m^o_\sigma \left[\log u_\sigma(w) - \sum_{\sigma'} m_{\sigma'}(t) \log u_{\sigma'}(w) \right] \\ &= \sum_\sigma m_\sigma(t) \left[\log u_\sigma(w) - 1\right] - \sum_\sigma m^o_\sigma \left[\log u_\sigma(w) - 1\right] \\ &\leq \phi(m(t)) - \phi(m^o),\end{aligned}$$

where the last line is because $w \in W(m(t))$. Now, Theorem 1 implies the last term is non-positive, so $\dot{D}_{\mathrm{KL}}\left(m^o || m(t)\right) \leq 0$ for all $t \geq 0$.

Any accumulation point m' of the trajectory $(m(t), t \geq 0)$ satisfies $\dot{D}_{\mathrm{KL}}\left(m^o || m'\right) = 0$ so $\phi(m') = \phi(m^o)$, by the above inequality, i.e., m' is a maximizer of ϕ.

As we could have taken m' as m^o, repeating the above for $m^o = m'$ yields $\lim_t \dot{D}_{\mathrm{KL}}(m' || m(t)) = 0$. Since m' is an accumulation point, we must have $\lim_t D_{\mathrm{KL}}(m' || m(t)) = 0$, so $m_\sigma(t) \to m'_\sigma$ for all σ, as $t \to \infty$. □

The condition in the theorem is necessary because extinct policies cannot re-emerge; thus, the set of initially non-extinct strategies must be rich enough to include an equilibrium. Otherwise, any stationary point reached will fail to be an equilibrium.

5.2.1 Joint learning and queueing dynamics. Here we assume the waiting delays $w(t)$ are not computed instantaneously given the current policy distribution $m(t)$, but evolve as fast as policies. In particular, their joint evolution is:

$$\dot{m}_\sigma(t) = m_\sigma \left(h_\sigma(u_\sigma(w(t))) - \sum_{\sigma'} m_{\sigma'} h_{\sigma'}(u_\sigma(w(t))) \right), \sigma \in \Pi, \tag{28}$$

$$\dot{w}_l(t) = \left[\frac{1}{b_l} \sum_\sigma \pi^\sigma_{ia} \alpha_{l,ia} x_\sigma(t) - 1 \right]^+_{w_l(t)}, \ l = 1, \ldots, L,$$

where $[z]^+_a = 0$ if $a = 0$ and $z < 0$; otherwise $[z]_a = z$. The second equation captures the queueing dynamics: the waiting delay of resource l grows proportionally to the instantaneous rate of requests for that resource and decreases at a constant rate of 1 (as time passes at a unit rate). Here, $x_\sigma(t)$ is the instantaneous policy rate of agents that use σ, and is defined by (20). Note that $x_\sigma(t)$ in general differs from $x_\sigma(m(t))$, as the resource constraints and the fluid queue condition now may not hold for all t.

For arbitrary but fixed $(m^o, w^o) \in \mathbb{R}^\Pi_+ \times \mathbb{R}^L_+$, define the function

$$\psi(m, w) = D_{\mathrm{KL}}\left(m^o || m\right) + \sum_l \frac{b_l}{2} (w_l - w^o_l)^2, \tag{29}$$

on $\mathbb{R}^\Pi_+ \times \mathbb{R}^L_+$. As the sum of two distances, it is always nonnegative and equals zero only if $(m, w) = (m^o, w^o)$.

Proposition 3. *The set of all primal-dual optimal pairs (w^o, m^o) of (25) is Lyapunov stable under the combined replicator and queueing dynamics (28).*

Proof of Proposition 3. Let m^o, w^o be as in the statement and define ψ as in (29). As in the proof of Proposition 2, it suffices to

show that $\psi(m(t), w(t))$ is increasing in t. To see this, note that

$$\dot{\psi}(m(t), w(t)) = \dot{D}_{\mathrm{KL}}\left(m^o || m(t)\right) + \sum_l b_l(w_l(t) - w_l^o)\dot{w}_l(t)$$
$$\leq \sum_\sigma m_\sigma(t)\left[\log u_\sigma(w(t)) - 1\right] - \sum_\sigma m_\sigma^o\left[\log u_\sigma(w(t)) - 1\right]$$
$$+ \sum_l (w_l(t) - w_l^o)\frac{\partial L}{\partial w_l}(w(t); m(t))$$
$$\leq L(w(t); m(t)) - L(w(t); m^o) + L(w^o; m(t)) - L(w(t); m(t)), \quad (30)$$

where $L(w; m)$ is the Lagrangian of (25), defined in (35), but with v omitted since all v terms vanish if m is a probability distribution.

The dual optimality of m^o implies $L(w^o; m(t)) \leq L(w^o; m^o)$, and the primal optimality of w^o implies $L(w(t); m^o) \geq L(w^o; m^o)$. Together, the two inequalities imply that the last expresion in (30) is nonpositive, so $\dot{\psi}(w(t), m(t)) \leq 0$. □

Lastly, note that if $t_{ia} = 0$ for all $i \in S, a \in A$ then the linear term in the optimization problem (23) vanishes. In this case, the average rewards $u_\sigma = r_\sigma x_\sigma / m_\sigma$ received by each agent using policy σ, for each σ, satisfy *proportionally fairness* [12]: any other feasible policy rates (x'_σ) result in average rewards $u'_\sigma = r_\sigma x'_\sigma / m'_\sigma$, for $\sigma \in P$, such that

$$\sum_\sigma m_\sigma \frac{u'_\sigma - u_\sigma}{u_\sigma} \leq 0.$$

The vanishing of the linear terms in (11) implies that the average reward for each agent at equilibrium matches the reward level that would arise under centralized coordination. (This follows because (11) is equivalent to the linear program that maximizes average rewards.) Consequently, the global asymptotic convergence under replicator dynamics in Proposition 2 translates here to:

Corollary 3. *When agents receive proportionally fair payoffs, the replicator dynamics converge to a socially optimal set of actions.*

Proof. Follows from Proposition 2. □

6 APPLICATION TO RIDEHAILING

In this section, we formulate a model of driver mobility in a ridehailing platform, where drivers act as agents.

The set of states, S, represents the different geographical regions where a driver can be located, and the set of actions is defined as:

$$A = \{\text{'wait'}, \text{'move to 1'}, \ldots, \text{'move to } |S|\text{'}\}.$$

The action 'wait' refers to a driver remaining in the current region until they are assigned a passenger originating from that region, whereas 'move to j' refers to a driver moving to region j without a passenger.

The MDP of each driver is defined by:

$$p_{ij}^a = \begin{cases} q_{ij}, & a = \text{'wait'}, \\ 1, & a = \text{'move to } j\text{'}, \\ 0, & \text{otherwise} \end{cases}, \quad t_{ia} = \begin{cases} \sum_j q_{ij} t_{ij}, & a = \text{'wait'}, \\ t_{ij}, & a = \text{'move to } j\text{'} \end{cases}$$

$$r_{ia} = \begin{cases} \sum_j q_{ij} c_i t_{ij}, & a = \text{'wait'}, \\ 0, & a = \text{'move to } j\text{'} \end{cases} \quad (31)$$

Here, c_i is the driver's compensation rate per unit time for transporting a passenger originating from region i. Under the action $a =$ 'move to j', the driver moves to j in t_{ij} time units without receiving compensation. Under $a =$ 'wait', the driver is assigned a passenger with destination j with probability q_{ij} and completes the journey in t_{ij} time units, receiving a total compensation of $c_i t_{ij}$.

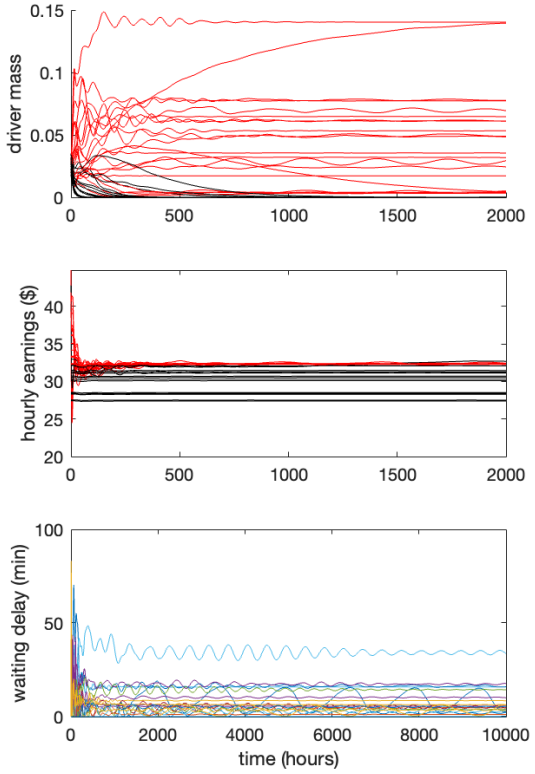

Figure 1: Ridehailing example based on NYC taxi data [6].

We also define S as the set of resources, where the i-th resource represents passengers originating from region i. The supply rate b_i for resource i is the demand for trips from region i, and $\alpha_{i,i\text{'wait'}} = 1$, and 0 otherwise, meaning that transporting a passenger from region i consumes one unit of resource i. (For simplicity, we assume that drivers are only assigned passengers from their current region, and that any excess passenger demand in a region is lost.)

Figure 1 illustrates the joint dynamics of learning and queuing (28) for an example based on NYC taxi data [6]. The NYC area is divided into $L = 64$ regions, with trip travel times (t_{ij}), passenger arrival rates (b_i), destination probabilities (q_{ij}), and driver compensation rates (c_i) derived from data collected for every Friday in March 2013, between 9 a.m. and 10 a.m.

The upper plot depicts the evolution of $m_\sigma(t)$ for each policy σ, assuming a uniform initial mass distribution. In the middle plot, the average reward $u_\sigma(w(t))$ is shown over time for each policy, with the best-performing policies shown in red. As indicated, the driver mass corresponding to these optimal policies (also shown in red in the upper plot) eventually dominates the entire driver population, rendering suboptimal policies (shown in black) extinct.

The lower plot depicts the waiting delay $w_i(t)$ in each region i over a longer time span. Long-lasting oscillations persist due to the presence of multiple equilibria.

7 SUMMARY

In this paper, we introduced a novel non-atomic model of resource competition where agents act according to a Markov Decision Process that is affected by resource congestion. In our model, action execution is associated with consuming resources from a common pool that get replenished with fixed rates, causing waiting captured by a fluid model. As a result, in this resource market, waiting plays the role of prices, discouraging agents to use actions that require congested resources. Agents subscribe to 'policies', i.e., a type of randomized cyclic behaviour, and collect the corresponding long-run average reward equal to the average reward of the cycle divided by the average cycle time. In an equilibrium, all cycles with positive agent mass have same average rewards, and convergence to the equilibrium occurs by agents switching to cycles that generate higher average revenue.

We show that equilibria correspond to optimal solutions of an extended Eisenberg-Gale program, that suggests the interpretation of our system as a market where delays play the role of prices. We introduce a new potential function formulation that allows us to study convergence of learning dynamics to the equilibria. This new formulation enables us to establish Lyapunov stability for a broad range of dynamics and prove global asymptotic stability for replicator dynamics. Furthermore, we demonstrate Lyapunov stability when agents follow replicator dynamics and queues adjust in similar time scales according to a natural tâtonnement mechanism.

Our research in this area started from our desire to model the effects of selfish agent optimization in practical applications consisting of closed systems where a fixed mass of circulating agents interacts continuously. There are many applications that fit into this category. For example, in ridehailing networks, a fixed population of drivers serves customers that arrive in the different nodes of the network, by circulating continuously according to different repositioning strategies. The equilibria of this ridehailing system are characterized by the solution of the extended Eisenberg-Gale program we mentioned earlier.

Finally, we note that, aside from the algorithm in Section 4, the results in this paper extend straightforwardly to the case of multiple agent types. However, these extensions are omitted here to simplify the notation.

APPENDIX

Proof of Lemma 2. Observe that (20), (22) are equivalent to the KKT conditions for problem (23), where w_l is the Lagrange multiplier for the capacity constraint of resource l. The optimal solution (x_σ) is unique since the objective is strictly concave with respect to x_σ whenever $m_\sigma > 0$; $\sum_{\sigma,i,a} \pi^\sigma_{ia} t_{ia} > 0$ implies $x_\sigma = 0$ when $m_\sigma = 0$. □

Proof of Lemma 3. The objective of the dual problem of $\max \phi(m)$ over probability distributions m is $\max [\phi(m) - v \sum_\sigma m_\sigma] + v$ where v is the dual variable. Replacing $\phi(m)$ by (25) yields

$$\max_{m\geq 0:\sum_\sigma m_\sigma=1} \left[\min_{w\geq 0} \sum_\sigma m_\sigma \left(\log u_\sigma(w) - 1 - v\right) + \sum_l b_l w_l \right] + v. \quad (32)$$

By viewing (m_σ) as Lagrange multipliers, we can interpret the maximization term as the dual of

$$\begin{aligned} \min\ & \sum_l b_l w_l \\ \text{s.t. } & \log u_\sigma(w) - 1 \leq v, \sigma \in \Pi, \\ \text{over } & w \geq 0. \end{aligned}$$

Replacing the maximization in (32) with the above problem yields (26). □

Proof of Proposition 1. To show that (x, w) is equilibrium we will use Theorem 1 but for the equivalent problem of (11),

$$\begin{aligned} \max\ & \log\left(\sum_\sigma r_\sigma x_\sigma\right) - \sum_{\sigma,i,a} \pi^\sigma_{ia} t_{ia} x_\sigma \quad (33) \\ \text{s.t. } & \sum_l \pi^\sigma_{ia} \alpha_{l,ia} x_\sigma \leq b_l,\ l = 1, \ldots, L, \\ \text{over } & x_\sigma \geq 0, \sigma \in \Pi, \end{aligned}$$

obtained by the change of variables in (21).

Now, let $(w, v), m$ be a primal and a dual solution, respectively of (26).

The complementary slackness conditions for the inequalities in (26) imply $e^{v+1} = u_\sigma(w)$ whenever $m_\sigma > 0$, and so

$$e^{v+1} = \sum_\sigma m_\sigma e^{v+1} = \sum_\sigma m_\sigma u_\sigma(w) = \sum_\sigma m_\sigma \frac{r_\sigma}{\tau_\sigma(w)} = \sum_\sigma x_\sigma r_\sigma,$$

by also using (20).

Thus, the inequalities (26) are rewritten as

$$\frac{r_\sigma}{\sum_{\sigma'} r_{\sigma'} x_{\sigma'}} \leq \tau_\sigma(w) = \sum_{i,a} \pi^\sigma_{ia} \left(t_{ia} + \sum_l \alpha_{l,ia} w_l \right), \quad (34)$$

with the inequality being an equality if $m_\sigma > 0$, equivalently $x_\sigma > 0$. This and the fact that (22) holds (since $w \in W(m)$), imply (x_σ) maximizes (33) and w is an optimal dual variable. Therefore, (x, w) is an equilibrium, by Theorem 1.

Conversely, let (x, w) be an equilibrium and $(x_\sigma)_\sigma$ any set of nonnegative coefficients such that (21) holds. (Such coefficients always exist, e.g., see [19].) By Theorem 1, x, w is a primal and dual optimal solution, respectively, of (11). The KKT conditions of (33) imply (34) holds, so

$$\frac{r_\sigma x_\sigma}{\sum_{\sigma'} r_{\sigma'} x_{\sigma'}} = \tau_\sigma(w) x_\sigma = m_\sigma.$$

Summation by parts over $\sigma \in \Pi$ yields $\sum_\sigma m_\sigma = 1$, and so m is a probability distribution.

Since (20), (22) hold, we have $w \in W(m)$ by the definition of the latter. By Lemma 2, w is an optimal solution of (25) and so it maximizes the Lagrangian of (26)

$$L(w, v; m) = \sum_\sigma m_\sigma \left(\log u_\sigma(w) - 1 - v\right) + \sum_l b_l w_l + v, \quad (35)$$

with respect to its first argument. Defining $v = \log\left(\sum_\sigma r_\sigma x_\sigma\right) - 1$, makes (w, v) a feasible point in (26) since

$$\log u_\sigma(w) - 1 = \log \frac{r_\sigma}{\tau_\sigma(w)} - 1 \leq \log \sum_{\sigma'} r_{\sigma'} x_{\sigma'} - 1 = v, \quad (36)$$

for all σ, where the inequality is due to (34). Finally, if $m_\sigma > 0$ then $x_\sigma > 0$, so the constraint in (36) is active, i.e., the complementary slackness condition of (26) hold. Therefore the KKT conditions are satisfied and so w, m is and optimal primal-dual pair. □

REFERENCES

[1] Martin J Beckmann, Charles B McGuire, and Christopher B Winsten. 1955. Studies in the Economics of Transportation. (1955).

[2] Benjamin Birnbaum, Nikhil R. Devanur, and Lin Xiao. 2011. Distributed algorithms via gradient descent for fisher markets. In *Proceedings of the 12th ACM Conference on Electronic Commerce* (San Jose, California, USA) *(EC '11)*. Association for Computing Machinery, New York, NY, USA, 127–136. https://doi.org/10.1145/1993574.1993594

[3] T. Chotibut, F. Falniowski, M. Misiurewicz, and G. Piliouras. 2020. The route to chaos in routing games: When is price of anarchy too optimistic? *Advances in Neural Information Processing Systems* 33 (2020), 766–777.

[4] John M. Danskin. 1966. The Theory of Max-Min, with Applications. *SIAM J. Appl. Math.* 14, 4 (1966), 641–664. http://www.jstor.org/stable/2946123

[5] Nikhil R. Devanur, Kamal Jain, Tung Mai, Vijay V. Vazirani, and Sadra Yazdanbod. 2016. New Convex Programs for Fisher's Market Model and its Generalizations. arXiv:1603.01257 [cs.GT] https://arxiv.org/abs/1603.01257

[6] Brian Donovan and Dan Work. 2016. New York City Taxi Trip Data (2010-2013). https://doi.org/10.13012/J8PN93H8

[7] Edmund Eisenberg and David Gale. 1959. Consensus of Subjective Probabilities: The Pari-Mutuel Method. *Ann. Math. Statist.* 30, 1 (03 1959), 165–168. https://doi.org/10.1214/aoms/1177706369

[8] Simon Fischer and Berthold Vöcking. 2004. On the Evolution of Selfish Routing. In *Algorithms – ESA 2004*, Susanne Albers and Tomasz Radzik (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 323–334.

[9] Josef Hofbauer and William Sandholm. 2009. Stable games and their dynamics. *Journal of Economic Theory* 144, 4 (2009), 1665–1693.e4. https://EconPapers.repec.org/RePEc:eee:jetheo:v:144:y:2009:i:4:p:1665-1693.e4

[10] Thomas Holding and Ioannis Lestas. 2021. Stability and Instability in Saddle Point Dynamics—Part I. *IEEE Trans. Automat. Control* 66, 7 (2021), 2933–2944. https://doi.org/10.1109/TAC.2020.3019375

[11] Kamal Jain and Vijay V. Vazirani. 2010. Eisenberg–Gale markets: Algorithms and game-theoretic properties. *Games and Economic Behavior* 70, 1 (2010), 84–106. https://doi.org/10.1016/j.geb.2008.11.011 Special Issue In Honor of Ehud Kalai.

[12] Frank Kelly. 1997. Charging and rate control for elastic traffic. *European Transactions on Telecommunications* 8, 1 (1997), 33–37. https://doi.org/10.1002/ett.4460080106

[13] F. P. Kelly. 1989. On a class of approximations for closed queueing networks. *Queueing Systems* 4, 1 (1989), 69–76. https://doi.org/10.1007/BF01150857

[14] Robert Kleinberg, Georgios Piliouras, and Eva Tardos. 2009. Multiplicative updates outperform generic no-regret learning in congestion games: extended abstract. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing* (Bethesda, MD, USA) *(STOC '09)*. Association for Computing Machinery, New York, NY, USA, 533–542. https://doi.org/10.1145/1536414.1536487

[15] Laurent Massoulié and James W. Roberts. 1999. Bandwidth sharing: objectives and algorithms. *IEEE INFOCOM '99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now (Cat. No.99CH36320)* 3 (1999), 1395–1403 vol.3. https://api.semanticscholar.org/CorpusID:16012348

[16] J. H. Nachbar. 1990. "Evolutionary" selection dynamics in games: Convergence and limit properties. *Int. J. Game Theory* 19, 1 (March 1990), 59–89. https://doi.org/10.1007/BF01753708

[17] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. 2007. *Algorithmic Game Theory*. Cambridge University Press, Cambridge. https://doi.org/DOI:10.1017/CBO9780511800481

[18] Georgios Piliouras and Fang-Yi Yu. 2023. Multi-agent performative prediction: From global stability and optimality to chaos. In *Proceedings of the 24th ACM Conference on Economics and Computation*. New York, NY, USA, 1047–1074. https://doi.org/10.1145/2623330.2623668

[19] Martin L Puterman. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

[20] Tim Roughgarden and Éva Tardos. 2002. How Bad is Selfish Routing? *J. ACM* 49, 2 (March 2002), 236–259. https://doi.org/10.1145/506147.506153

[21] Paul Schweitzer. 1981. Approxiamate analysis of multiclass closed networks of queues. *J. ACM* 29, 2 (1981).

[22] V. I. Shmyrev. 2009. An algorithm for finding equilibrium in the linear exchange model with fixed budgets. *Journal of Applied and Industrial Mathematics* 3, 4 (2009), 505–518. https://doi.org/10.1134/S1990478909040097

[23] J G Wardrop. 1952. SOME THEORETICAL ASPECTS OF ROAD TRAFFIC RESEARCH. *Proceedings of the Institution of Civil Engineers* 1, 3 (1952), 325–362. https://doi.org/10.1680/ipeds.1952.11259

[24] Li Zhang. 2011. Proportional response dynamics in the Fisher market. *Theoretical Computer Science* 412, 24 (2011), 2691–2698. https://doi.org/10.1016/j.tcs.2010.06.021 Selected Papers from 36th International Colloquium on Automata, Languages and Programming (ICALP 2009).