

# The Nash Equilibrium with Inertia in Population Games

Basilio Gentile, Dario Paccagnan, Bolutife Ogunsula, and John Lygeros

**Abstract**—In the traditional game-theoretic set up, where agents select actions and experience corresponding utilities, an equilibrium is a configuration where no agent can improve their utility by unilaterally switching to a different action. In this work, we introduce the novel notion of inertial Nash equilibrium to account for the fact that, in many practical situations, action changes do not come for free. Specifically, we consider a population game and introduce the coefficients  $c_{ij}$  describing the cost an agent incurs by switching from action  $i$  to action  $j$ . We define an inertial Nash equilibrium as a distribution over the action space where no agent benefits in moving to a different action, while taking into account the cost of this change. First, we show that the set of inertial Nash equilibria contains all the Nash equilibria, but is in general not convex. Second, we argue that classical algorithms for computing Nash equilibria cannot be used in the presence of switching costs. We then propose a natural better-response dynamics and prove its convergence to an inertial Nash equilibrium. We apply our results to predict the drivers' distribution of an on-demand ride-hailing platform.

## I. INTRODUCTION

Game theory has originated as a set of tools to model and describe the interaction of multiple decision makers, or agents. The goal is typically to determine whether decision makers will come to some form of equilibrium, the most common of which is the Nash equilibrium. Informally, a set of strategies constitutes a Nash equilibrium if no agent benefits by unilaterally deviating from the current action, while the other agents stay put. This notion of equilibrium has found countless applications, among others to energy systems [1], transmission networks [2], commodity markets [3], traffic flow [4], and mechanism design [5].

While the original definition of Nash equilibrium does not account for the cost incurred by agents when moving to a different action, in practical situations decision makers often incur a physical, psychological, or monetary cost for such deviation. This is the case, for example, when relocating to a new neighbourhood [6], or when switching financial strategy in the stock market [7]. When the decision makers are humans, the psychological resistance to change has been well documented and studied at the professional and organizational level [8] as well as at the individual and private level [9], or at the customer level [10].

To take into account such phenomena, we introduce the novel concept of *inertial Nash equilibrium*. Specifically, we consider a setup where a large number of agents choose among  $n$  common actions. Agents selecting a given action receive a utility that depends only on the agents' distribution over the action space, in the same spirit of population games [11]. In this context, a Nash equilibrium consists in an agent distribution over the action space for which every utilized action yields maximum utility. The same concept was proposed in

the seminal work of Wardrop for a route-choice game in road traffic networks [4]. We extend this framework and model the cost incurred by any agent when moving from action  $i$  to action  $j$  with the non-negative coefficients  $c_{ij}$ . We define an *inertial Nash equilibrium* as a distribution over the action space where no agent has any incentive to unilaterally change action, where the quality of an alternative action is measured by its *net utility*, i.e., the corresponding utility *minus* the cost of the action change.

We show that introducing such costs leads to a larger set of equilibria that is in general not convex, even if the set of Nash equilibria without switching costs is so. We argue that classical algorithms to compute a Nash equilibrium are not suitable for computing inertial Nash equilibria, because i) they may not terminate even if already at an inertial Nash equilibrium, and ii) their execution is not compatible with the agents' rationality assumption, as agents might be required to perform a detrimental move. To overcome these issues, we propose an algorithm based on better-response dynamics, where agents switch action only if it is to their advantage when factoring the cost of such change.

**Contributions.** Our main contributions are as follows.

- i) We introduce the notion of inertial Nash equilibrium (Definition 2) and position it in the context of the existing literature, notably in relation to population games [11] and more specifically migration equilibria [12].
- ii) We show that the set of inertial Nash equilibria can be equivalently characterized through a variational inequality (Theorem 1) and we prove a strong negative result: the operator that arises in the resulting variational inequality is non-monotone in all the meaningful instances of the inertial Nash equilibrium problem (Theorem 2). This implies that existing algorithms for computing equilibria based on the solution of variational inequalities are in general not suitable for computing an inertial Nash equilibrium.
- iii) We propose and analyse a novel algorithm and prove its convergence to an inertial Nash equilibrium under weak assumptions (Theorem 3).

**Organization.** In Section II we introduce the notion of inertial Nash equilibrium, and show its non-uniqueness as well as the non-convexity of the equilibrium set. A comparison with related works is presented in Section II-C. In Section III we reformulate the inertial Nash equilibrium problem as a variational inequality, study the monotonicity properties of the corresponding operator (more precisely, the lack thereof), and present the issues associated with the use of existing algorithms. In Section IV we propose a modified best-response dynamics that provably converges to an inertial Nash equilibrium. Extensions of the model are presented in Section V.

This work was supported by the European Commission project DYMASOS (FP7-ICT 611281), and by the SNSF Grant #P2EZP2-181618.

In Section VI we validate our model with a numerical study of area coverage for on demand ride-hailing in Hong Kong. Appendix A provides background material, while all the proofs are reported in Appendix B.

**Notation.** The space of  $n$ -dimensional real vectors is denoted with  $\mathbb{R}^n$ , while  $\mathbb{R}_{\geq 0}^n$  is the space of non-negative  $n$ -dimensional real vectors and  $\mathbb{R}_{> 0}^n$  is the space of strictly positive  $n$ -dimensional real vectors. The symbol  $\mathbb{1}_n$  indicates the  $n$ -dimensional vector of unit entries, whereas  $\mathbb{0}_n$  is the  $n$ -dimensional vector of zero entries. If  $x, y \in \mathbb{R}^n$ , the notation  $x \geq y$  indicates that  $x_j \geq y_j$  for all  $j \in \{1, \dots, n\}$ . The vector  $e_i$  denotes the  $i^{\text{th}}$  vector of the canonical basis. Given  $A \in \mathbb{R}^{n \times n}$ ,  $A \succ 0$  ( $\succeq 0$ ) if and only if  $x^\top A x = \frac{1}{2} x^\top (A + A^\top) x > 0$  ( $\geq 0$ ), for all  $x \neq \mathbb{0}_n$ .  $\text{blkdiag}(A_1, \dots, A_M)$  is the block diagonal matrix with blocks  $A_1, \dots, A_M$ .  $\|A\|$  is the induced 2-norm on  $A$ . Given  $g(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  we define  $\nabla_x g(x) \in \mathbb{R}^{n \times m}$  with  $[\nabla_x g(x)]_{i,j} := \frac{\partial g_j(x)}{\partial x^i}$ . If  $n = m = 1$ , we use  $g'(x)$  to denote the derivative of  $g$  at the point  $x$ .  $I_n$  denotes the  $n \times n$  identity matrix.  $\text{Proj}_{\mathcal{X}}[x]$  is the Euclidean projection of the vector  $x$  onto a closed and convex set  $\mathcal{X}$ .

## II. INERTIAL NASH EQUILIBRIUM: DEFINITION AND EXAMPLES

### A. Definition of Inertial Equilibrium

We consider a large number of competing agents with a finite set of common actions  $\{1, \dots, n\}$ . For selecting action  $i \in \{1, \dots, n\}$ , an agent receives a utility  $u_i(x)$ , where  $x = [x_1, \dots, x_n]$ , and  $x_i$  denotes the fraction of agents selecting action  $i$ . Observe that, with the introduction of the utility functions  $u_i : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}$ , we are implicitly assuming that the utility received by playing action  $i$  only depends on the distribution of the agents, and not on which agent selected which action, a modelling assumption typically employed in population games [11]. Within this framework, a Nash equilibrium is a distribution over the action space where no agent has any incentive in deviating to a different action. This requirement can be formalized by introducing the unit simplex<sup>1</sup> in dimension  $n$ , denoted with  $\mathcal{S}$ , and its relative interior  $\mathcal{S}_+$

$$\begin{aligned} \mathcal{S} &:= \{x \in \mathbb{R}^n \text{ s.t. } x \geq 0, \mathbb{1}_n^\top x = 1\}, \\ \mathcal{S}_+ &:= \{x \in \mathbb{R}^n \text{ s.t. } x > 0, \mathbb{1}_n^\top x = 1\}. \end{aligned}$$

**Definition 1** (Nash equilibrium, [4]). *Given  $n$  utilities  $\{u_i\}_{i=1}^n$  with  $u_i : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}$ , the vector  $\bar{x} \in \mathcal{S}$  is a Nash equilibrium if*

$$\bar{x}_i > 0 \implies u_i(\bar{x}) \geq u_j(\bar{x}), \quad \forall i, j \in \{1, \dots, n\}. \quad (1)$$

Despite being widely used in the applications, Definition 1 does not account for the cost associated with an action switch. We extend the previous model by introducing the non-negative coefficients  $c_{ij}$  to represent the cost experienced by any agent when moving from action  $i$  to  $j$ . We then define an inertial Nash equilibrium as a distribution over the action space where no agent can benefit by moving to a different action, while taking into account the cost of such change.

**Definition 2** (Inertial Nash equilibrium). *Given  $n$  utilities  $\{u_i\}_{i=1}^n$ ,  $u_i : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}$ ,  $n^2$  non-negative switching costs  $\{c_{ij}\}_{i,j=1}^n$ , the vector  $\bar{x} \in \mathcal{S}$  is an Inertial Nash equilibrium if*

$$\bar{x}_i > 0 \implies u_i(\bar{x}) \geq u_j(\bar{x}) - c_{ij}, \quad \forall i, j \in \{1, \dots, n\}. \quad (2)$$

In the remainder of this manuscript we focus on problems where there is no cost for staying put, as formalized next.

**Standing assumption.** *The switching costs satisfy  $c_{ii} = 0$  for all  $i \in \{1, \dots, n\}$ .*

Observe that conditions (1) and (2) do not impose any constraint on actions that are not currently selected by any agent (i.e., those with  $\bar{x}_i = 0$ ). In other words, the utility of one such action can be arbitrarily low, and the configuration  $\bar{x}$  still be an equilibrium. Despite being a natural extension to the traditional notions of equilibrium in game theory, to the best of our knowledge, Definition 2 is novel. Its relevance stems from the observation that the coefficients  $c_{ij}$  can model different and common phenomena, such as:

- the tendency of agents to adhere to their habits, or their reluctance to try something different;
- actual costs or fees that agents incur for switching action;
- the lack of accurate information about other options.

In the following, we provide two examples of problems that can be captured within this framework.

**On demand ride-hailing.** Ride-hailing systems are platforms that allow customers to travel from a given origin to a desired destination, typically within the same city. Examples include taxi companies as well as platforms such as Uber, Lyft or Didi. In our framework, the drivers correspond to agents and geographical locations to available actions. Each utility describes the profitability of a given location, which depends on the arrival rate of customers in that location, and on the fraction of vehicles available in that same location. The cost (fuel and time) that a driver incurs while moving between two different physical locations is captured by  $c_{ij}$ . Such model can predict how drivers distribute themselves over the city.

**Task assignment in server network.** We are given a finite number of geographically dispersed servers represented with nodes, and connected through a network. Each server corresponds to an action  $i \in \{1, \dots, n\}$ . A large number of agents has a list of jobs that originates in various nodes on the network and wishes to execute this list as swiftly as possible. The speed at which each server can process a job depends on the load on the server and is captured by  $u_i(x_i)$ . Moving a job between server  $i$  and  $j$  requires an amount of time and resources captured by  $c_{ij}$ . This model can predict how agents distribute their jobs over the set of servers.

We note that the set of inertial Nash equilibria contains the set of Nash equilibria, due to the non negativity of  $c_{ij}$ .

**Lemma 1.** *Every Nash equilibrium is an inertial Nash equilibrium.*

The proof follows from Definition 1 and 2, since condition (1) implies condition (2), as  $c_{ij} \geq 0$  for all  $i, j \in \{1, \dots, n\}$ .

<sup>1</sup>The formulation with unitary mass and the corresponding results seamlessly generalise to agents of combined mass  $\gamma > 0$ .

In the following we refer to an (inertial) Nash equilibrium as just an (inertial) equilibrium.

### B. Non-uniqueness and non-convexity of the equilibrium set

The following example shows that the set of inertial equilibria is in general neither convex, nor a singleton. This will pose significant algorithmic challenges, as discussed in Section III.

**Example 1.** Let  $n = 3$ , and consider utilities and switching costs of the form

$$\begin{aligned} u_1(x) &= 1.2 - x_1 \\ u_2(x) &= 1.2 - x_2 \\ u_3(x) &= 1 - x_3, \end{aligned} \quad C = \begin{bmatrix} 0 & 0.2 & 0.3 \\ 1 & 0 & 0.8 \\ 0.1 & 1.2 & 0 \end{bmatrix},$$

where the entry  $(i, j)$  of  $C$  equals  $c_{ij}$ . Note that  $x_3 = 1 - x_1 - x_2$ . The equilibrium conditions (2) then become

$$x_1 > 0 \quad \Rightarrow \quad x_2 \geq x_1 - 0.2 \quad (3a)$$

$$x_1 > 0 \quad \Rightarrow \quad x_2 \leq -2x_1 + 1.5 \quad (3b)$$

$$x_2 > 0 \quad \Rightarrow \quad x_2 \leq x_1 + 1 \quad (3c)$$

$$x_2 > 0 \quad \Rightarrow \quad x_2 \leq -0.5x_1 + 1 \quad (3d)$$

$$x_3 > 0 \quad \Rightarrow \quad x_2 \geq -2x_1 + 1.1 \quad (3e)$$

$$x_3 > 0 \quad \Rightarrow \quad 2x_2 \geq -x_1, \quad (3f)$$

where inequalities (3c), (3d), (3f) are already implied by  $x \in \mathcal{S}$ . We color the remaining three inequalities similarly to Figure 1, which reports the solution to (3) (i.e., the inertial equilibrium set) in gray.

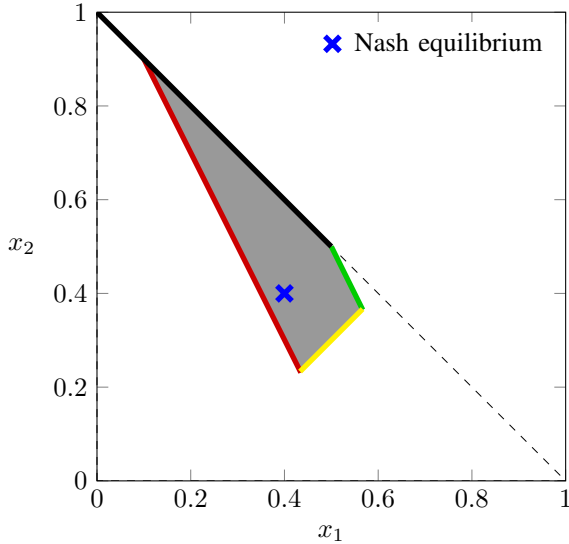


Fig. 1: The shaded region, including the thick red, yellow, green, and black lines, represents the inertial Nash equilibrium set for Example 1 projected on the plane  $(x_1, x_2)$ . The component  $x_3$  can be reconstructed from  $x_3 = 1 - x_1 - x_2$ . The dashed line represents the simplex boundary, while the yellow, green and red lines describe the inequalities in (3). The blue point is the unique Nash equilibrium  $\bar{x} = [0.4, 0.4, 0.2]$ , which satisfies condition (1).

We note that the inertial equilibrium set is not a singleton. The lack of uniqueness is due to the positivity of the

coefficients  $c_{ij}$ . Indeed, if  $c_{ij} = 0$  for all  $i, j$ , then the inertial equilibrium set coincides with the equilibrium set of Definition 1, which is a singleton marked in blue in Figure 1. Moreover, the inertial equilibrium set is not convex. This is due to the line joining the point  $(0.1, 0.9)$  to  $(0, 1)$  in Figure 1. The points on this segment belong to the inertial equilibrium set even though they do not satisfy  $x_2 \geq -2x_1 + 1.1$ . This is because (3e) is enforced only when  $x_3 > 0$ , whereas  $x_3 = 0$  on the considered segment. The observed non-convexity of the solution set is, in a sense, structural. To see this, note that, by Definition 2, a point  $x \in \mathcal{S}_+$  is an inertial equilibrium if and only if it lies at the intersection of inequality constraints of the form  $u_j(x) - c_{ij} - u_i(x) \leq 0$ ; these might be non convex, even if we restrict attention to convex or concave utility functions.

### C. Related Work

The notion of inertial equilibrium is, to the best of our knowledge, novel, due to the presence of the switching costs  $c_{ij}$ . A related line of works comes from *population games* [13]. Here the focal point is the analysis and design of (continuous-time) agent dynamics that achieve an equilibrium in the sense of Definition 1. A particular class of dynamics is imitation dynamics. These are reminiscent of the discrete-time Algorithm 2 below, as agents move to more attractive actions. Different works provide local [11], [14] and global [15], [16] convergence guarantees. Rather than delving into the vast literature of population games, we observe that in all of the works there is no switching cost, i.e.,  $c_{ij} = 0$ . Thus, the literature of population games study the problem of finding an equilibrium in the sense of Definition 1 and not an equilibrium in the sense of Definition 2, which is the focus here. Finally, we note that [11] and references therein provide convergence results to an equilibrium set, whereas we provide convergence to a point in the inertial equilibrium set.

A more closely related equilibrium concept was proposed in the study of migration models in the seminal works [12], [17], [18] by Nagurney. These works introduce the notion of migration equilibrium, in a way that resembles Definition 2, but with a number of important differences. First, the problem formulation is different. In the migration equilibrium problem we are given a fixed initial distribution  $x^0 \in \mathcal{S}$ , with  $x_j^0$  representing the fraction of agents residing at a physical location  $j$ . These agents receive utility  $u_j(x^0)$ . The initial distribution  $x^0$  is transformed into the final distribution  $x^1 \in \mathcal{S}$ , which is a function of the migrations  $(f_{ij})_{i,j=1}^n$  (the decision variables). Each migration comes with a migration cost  $c_{ij}(f_{ij})$  which is a function of the number  $f_{ij}$  of agents migrating. A migration equilibrium consists of a set of migrations  $(f_{ij})_{i,j=1}^n$  such that, considering the fixed initial utilities  $u(x^0)$ , the migration costs  $c_{ij}(f_{ij})$  and the final utilities  $u(x^1)$ , no other set of migrations is more convenient. Second, while the better-response algorithm we will introduce in Section IV can be interpreted as the natural dynamics of the agents seeking an equilibrium, this is not the case for the algorithms proposed to find a migration equilibrium, which are instead VI algorithms to be carried out offline.

### III. VARIATIONAL INEQUALITY REFORMULATION

In this section we first recall that the set of equilibria defined by (1) can be described as the solution of a certain variational inequality. We then show that a similar result holds for the inertial equilibrium set of (2). While the former equivalence is known, the latter connection is novel and requires the careful definition of the variational inequality operator. The interest in connecting the inertial equilibrium problem with the theory of variational inequalities stems from the possibility of inheriting readily available results, such as existence of the solution, properties of the solution set, and algorithmic convergence. Basic properties and results from the theory of variational inequalities used in this manuscript are summarized in Appendix A.

**Definition 3** (Variational inequality). *Consider a set  $\mathcal{X} \subseteq \mathbb{R}^n$  and an operator  $F : \mathcal{X} \rightarrow \mathbb{R}^n$ . A point  $\bar{x} \in \mathcal{X}$  is a solution of the variational inequality  $\text{VI}(\mathcal{X}, F)$  if*

$$F(\bar{x})^\top (x - \bar{x}) \geq 0, \quad \forall x \in \mathcal{X}.$$

The variational inequality problem was first introduced in infinite dimensional spaces in [19], while the finite-dimensional problem in Definition 3 was identified and studied for the first time in [20]. The monograph [21] includes a wide range of results on VI, amongst which their connection to Nash equilibria.

**Proposition 1** (Equilibria as VI solutions, [13, Thm 2.3.2]). *A point  $\bar{x} \in \mathcal{S}$  is an equilibrium if and only if it is a solution of  $\text{VI}(\mathcal{S}, -u)$ , where  $u(x) := [u_i(x)]_{i=1}^n$ .*

The following theorem shows that inertial equilibria can also be described by suitable variational inequalities.

**Theorem 1** (Inertial equilibria as VI solutions). *A point  $\bar{x} \in \mathcal{S}$  is an inertial equilibrium if and only if it is a solution of  $\text{VI}(\mathcal{S}, F)$ , where*

$$\begin{aligned} F(x) &:= [F_i(x)]_{i=1}^n, \\ F_i(x) &:= \max_{j \in \{1, \dots, n\}} (u_j(x) - u_i(x) - c_{ij}). \end{aligned} \quad (4)$$

*If the utilities are continuous, the existence of an inertial equilibrium is guaranteed.*

#### A. Lack of monotonicity

If the operator  $F$  in  $\text{VI}(\mathcal{S}, F)$  is monotone (see Definition 6 in Appendix A), an inertial equilibrium can be computed efficiently using one of the many algorithms available in the literature of variational inequalities, see [21, Chapter 12]. On the contrary, if this is not the case, the problem is known to be intractable in general. Since the inertial equilibrium set of Figure 1 is not convex, the corresponding variational inequality operator  $F$  cannot be monotone (see Proposition 7 in Appendix A). The question is whether this observation extends to more general settings. In the following we provide a strong negative result showing that the variational inequality operator is *non monotone* in many meaningful instances of the inertial equilibrium problem.

**Theorem 2** ( $F$  is not monotone). *Assume that for all  $i \in \{1, \dots, n\}$  the function  $u_i$  is Lipschitz and that  $\nabla_{x_i} u_i(x) < 0$  for all  $x \in \mathcal{S}$ . If there exists a point  $\hat{x} \in \mathcal{S}$  which is not an inertial equilibrium, then  $F$  is not monotone in  $\mathcal{S}$ .*

The theorem certifies that either every point of the simplex is an equilibrium, or  $F$  is not monotone and consequently the variational inequality problem is hard. The only technical assumption is that  $\nabla_{x_i} u_i(x) < 0$ . We observe that this is the situation for many applications; indeed the condition implies that  $u_i(x)$  decreases if the number of agents on action  $i$  increases, as commonly assumed in congestion problems. Moreover, the condition can be further weakened, as for the proof it suffices that  $\nabla_{x_{i^*}} u_{i^*}(x^*) < 0$ , only for a specific  $x^*$  and  $i^*$  defined in Appendix B.

We conclude this section by pointing out that Example 1 satisfies the conditions of Theorem 2. The lack of monotonicity of the corresponding operator  $F$  is confirmed by the fact that  $\nabla_x F(x)$  is not positive semidefinite for all  $x \in \mathcal{S}$  (a condition equivalent to monotonicity, see Proposition 6 in Appendix A). Indeed, there are points where  $\nabla_x F(x) + \nabla_x F(x)^\top$  is indefinite, e.g.,  $\tilde{x} = [0.2, 0.2, 0.6]$ , where

$$\nabla_x F(\tilde{x}) + \nabla_x F(\tilde{x})^\top = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 2 \end{bmatrix}.$$

#### B. Three drawbacks of existing algorithms

Lemma 1 ensures that any equilibrium is an inertial equilibrium. Thus, one might be tempted to use an algorithm for computing an equilibrium to determine an inertial equilibrium. Unfortunately, a number of difficulties make this approach impractical. In this section we describe *one* such algorithm and highlight its drawbacks in the computation of an inertial equilibrium, which generalise to other algorithms that converge to an equilibrium. We consider the projection algorithm [21, Alg. 12.1.1] for the solution of  $\text{VI}(\mathcal{S}, -u)$ , where  $x(k)$  indicates the iterate  $k$  of the algorithm. Note that the projection step necessitates the presence of a central operator.

---

#### Algorithm 1 Projection algorithm

---

**Initialization:**  $\rho > 0, k = 0, x(0) \in \mathcal{S}$   
**Iterate:**  $x(k+1) = \text{Proj}_{\mathcal{S}} [x(k) + \rho u(x(k))]$   
 $k \leftarrow k+1$

---

**Proposition 2.** *If  $u_i$  is  $L$ -Lipschitz for all  $i$ ,  $\rho \leq 2/L$ , and if there exists a concave function  $\theta : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $\nabla_x \theta(x) = u(x)$  for all  $x \in \mathcal{S}$ , then Algorithm 1 converges to an equilibrium, and thus an inertial equilibrium.<sup>2</sup>*

<sup>2</sup> For this proposition to hold, we have to assume the existence of a concave function  $\theta$  whose gradient matches  $u(x)$ . One such case is when the utility function  $u_i$  depends only on the number of agents on action  $i$ , i.e.  $u_i(x) = u_i(x_i)$  for all  $i$ , and is decreasing. This case covers a wide range of applications. If no  $\theta$  whose gradient matches  $u(x)$  exists, but  $-u$  is monotone, one can resort to a different algorithm such as the extra-gradient algorithm [21, Thm. 12.1.11]. Finally, observe that if  $-u$  is strongly monotone (see [21, Def. 2.3.1]), the projection algorithm converges without requiring the existence of  $\theta(x)$ , see [21, Alg. 12.1.1].

### Three fundamental shortcomings

In the following we analyse the behaviour of Algorithm 1 on Example 1, and use it to highlight three fundamental shortcomings of this approach. We begin by observing that  $\bar{x} = [\bar{x}_1, \bar{x}_2, \bar{x}_3] = [0.4, 0.4, 0.2]$  is an equilibrium, as it solves  $\text{VI}(\mathcal{S}, -u)$ , since for all  $x \in \mathcal{S}$

$$\begin{bmatrix} -u_1(\bar{x}_1) \\ -u_2(\bar{x}_2) \\ -u_3(\bar{x}_3) \end{bmatrix}^\top \left( \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} \right) = \begin{bmatrix} -0.8 \\ -0.8 \\ -0.8 \end{bmatrix}^\top \left( \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} \right) = 0.$$

Additionally,  $[\bar{x}_1, \bar{x}_2, \bar{x}_3] = [0.4, 0.4, 0.2]$  is the *unique* solution of  $\text{VI}(\mathcal{S}, -u)$  and thus the unique equilibrium (see [21, Thm. 2.3.3]). This is consistent with Lemma 1 and Figure 1 where the equilibrium point  $\bar{x}$  belongs to the inertial equilibrium set. Thanks to Proposition 2, Algorithm 1 converges to  $\bar{x}$  ( $L = 1$  for the utilities in (3), so that we have to select  $\rho < 2$ ). With the choice of  $\rho = 1$ , it is immediate to verify that Algorithm 1 converges in one iteration for any initial condition  $x(0)$ .

In the following we consider two cases: i) the case in which  $x(0)$  is neither an inertial equilibrium nor an equilibrium; ii) the case in which  $x(0)$  is an inertial equilibrium, but not an equilibrium. Case i): consider  $x(0) = [0.4, 0.2, 0.4]$ . The point  $x(0)$  is not an inertial equilibrium (and thus not an equilibrium), because  $x_3(0) > 0$  and  $u_3(x(0)) = 1 - 0.4 = 0.6 < 0.7 = 0.8 - 0.1 = u_1(x(0)) - c_{31}$ . The first iteration of Algorithm 1 amounts to a mass of 0.2 being moved from action  $i = 3$  to action  $i = 2$ . Nevertheless, we observe that agents selecting action  $i = 3$  are not interested in moving to action  $i = 2$ . Indeed  $u_3(x(0)) = 0.6 \geq -0.2 = u_2(x(0)) - c_{32}$ , so the switch from  $i = 3$  to  $i = 2$  is detrimental for the agents performing it. Case ii): Consider  $x(0) = [0.4, 0.3, 0.3]$ , and note that  $x(0)$  is already an inertial equilibrium. Nonetheless, the first iteration of Algorithm 1 forces a mass of 0.1 to move from action 3 to 2.

The drawbacks of Algorithm 1 are summarized next:

- i) Agents are forced to switch action even when such switch is detrimental to their well being.
- ii) Agents are forced to switch action even if already at an inertial equilibrium.
- iii) The projection step necessitates the presence of a central operator. Such operator requires information not only on the utilities  $u_i(x(k))$  for all  $i$ , but also on  $x(k)$ .

In the next section we overcome these issues and present a natural dynamics that i) provably converges to an inertial equilibrium, ii) respects the agent's strategic nature, and iii) requires limited coordination.

## IV. A BETTER-RESPONSE ALGORITHM

We begin by introducing the definition of the envy set.

**Definition 4** (Envy set). *Given  $x \in \mathcal{S}$ , for each  $i$  such that  $x_i > 0$ , we define the envy set of  $i$  as*

$$\mathcal{E}_i^{\text{out}}(x) := \{j \in \{1, \dots, n\} \text{ s.t. } u_i(x) < u_j(x) - c_{ij}\},$$

whereas for  $i$  such that  $x_i = 0$ , we define  $\mathcal{E}_i^{\text{out}}(x) = \emptyset$ .

Informally, the envy set  $\mathcal{E}_i^{\text{out}}(x)$  contains all the actions  $j$  to which agents currently selecting action  $i$  would rather move

to. The following fact immediately follows from Definition 4 and Definition 2 of inertial equilibrium.

**Proposition 3.** *A point  $\bar{x} \in \mathcal{S}$  is an inertial equilibrium if and only if  $\bar{x} \in \mathcal{S}$  and  $\mathcal{E}_i^{\text{out}}(\bar{x}) = \emptyset$ , for all  $i \in \{1, \dots, n\}$ .*

The proposed Algorithm 2 involves a single, intuitive step. At iteration  $k$ , let  $x(k) \in \mathcal{S}$  denote the distribution of the agents on the resources. For every action  $i$ , a mass  $x_{i \rightarrow j}(k) \in [0, x_i(k)]$  is moved from action  $i$  to some other action  $j \in \mathcal{E}_i^{\text{out}}(x(k))$ , that is, the movement takes place only if the alternative action  $j$  is attractive for agents currently selecting action  $i$ . This simple dynamics is described in Algorithm 2, where we denote with  $u_i(k) = u_i(x(k))$ ,  $\mathcal{E}_i^{\text{out}}(k) = \mathcal{E}_i^{\text{out}}(x(k))$  for brevity.

---

### Algorithm 2 Better-response algorithm

---

**Initialization:**  $k = 0, x(0) \in \mathcal{S}$   
**Iterate:**  $\Delta x(k) \leftarrow 0$   
**repeat** for all  $i, j \in \mathcal{E}_i^{\text{out}}(k)$   
    choose  $x_{i \rightarrow j}(k) \in [0, x_i(k)]$   
     $\Delta x_i(k) \leftarrow \Delta x_i(k) - x_{i \rightarrow j}(k)$ ,  
     $\Delta x_j(k) \leftarrow \Delta x_j(k) + x_{i \rightarrow j}(k)$ ,  
**end repeat**  
 $x(k+1) \leftarrow x(k) + \Delta x(k)$   
 $k \leftarrow k+1$

---

The agents' dynamics presented in Algorithm 2 is fully specified once we define the mass  $x_{i \rightarrow j}(k)$  moving from action  $i$  to  $j \in \mathcal{E}_i^{\text{out}}(k)$  as a function of  $x_i(k)$  and  $x_j(k)$ . At this stage we rather not give a particular expression to  $x_{i \rightarrow j}(k)$ , as the convergence of Algorithm 2 is guaranteed under very weak conditions and different choices of  $x_{i \rightarrow j}(k)$ . A possible modelling assumption sees agents moving from a less attractive action  $i$  to a more favourable action  $j \in \mathcal{E}_i^{\text{out}}(k)$  *independently* from the value of the utility  $u_j(k)$ . For instance, this can be achieved by setting  $x_{i \rightarrow j}(k) = \beta x_i(k)$  with  $\beta > 0$ . A different modelling assumption entails agents being responsive to the level of the utility  $u_j(k)$  over all  $j \in \mathcal{E}_i^{\text{out}}(k)$ , and thus redistributing themselves based on the perceived gain. Both these cases (and many more) are covered by Theorem 3.

We observe that Algorithm 2 does not present any of the issues encountered with the use of Algorithm 1. First, agents switch action only if the switch is convenient. Second, no agent moves if the current allocation is an inertial equilibrium. Third, there is no need for a central operator, and each agent requires information only regarding the other actions' utilities  $u(x(k))$ . As a consequence, Algorithm 2 can be interpreted as the *natural dynamics* of agents switching to a more favourable action whenever one is available. Finally, agents are not limited to moving to the best alternative action (as in *best-response dynamics*), but can instead choose any action providing a better net utility (hence the term *better-response dynamics*).

**Theorem 3** (Convergence of Algorithm 2). *Assume that*

- *for each  $i \in \{1, \dots, n\}$  the utility  $u_i$  depends only on  $x_i$ , that  $u_i$  is non-increasing and  $L$ -Lipschitz.*

- there exists  $c_{\min} > 0$  such that  $c_{ij} \geq c_{\min}$  for all  $i \neq j$  with  $i, j \in \{1, \dots, n\}$ .
- there exist  $0 < \tau \leq 1$ , and  $\varepsilon > 0$  such that at each iteration  $k \in \mathbb{N}$ ,  $x_{i \rightarrow j}(k) \geq 0$  for all  $i \in \{1, \dots, n\}$ ,  $j \in \mathcal{E}_i^{\text{out}}(x_k)$ , and

$$\tau x_i(k) \leq \sum_{j \in \mathcal{E}_i^{\text{out}}(k)} x_{i \rightarrow j}(k) \leq x_i(k), \quad i \in \{1, \dots, n\}, \quad (5a)$$

$$\sum_{i: j \in \mathcal{E}_i^{\text{out}}(k)} x_{i \rightarrow j}(k) \leq \frac{c_{\min}}{L} - \varepsilon, \quad j \in \{1, \dots, n\}. \quad (5b)$$

Then  $x(k)$  in Algorithm 2 converges to an inertial equilibrium  $\bar{x}$ . If additionally  $\bar{x} \in \mathcal{S}_+$ , then the algorithm terminates in a finite number of steps.

The first assumption is typical of many congestion-like problems. The second assumption is technical, and requires the switching costs between different actions to be strictly positive. With respect to the third assumption, the requirement on the right hand side of (5a) together with the condition  $x_{i \rightarrow j}(k) \geq 0$  for all  $i \in \{1, \dots, n\}$ ,  $j \in \mathcal{E}_i^{\text{out}}(x_k)$ , is needed to ensure that  $x(k)$  remains in the simplex. Thus, the only non-trivial constraint imposed on  $x_{i \rightarrow j}(k)$  is that on the left hand side of (5a), and that of (5b); these are discussed in detail in Remark 1 below. Finally, we note that the proof of Theorem 3 does not require the agents to move synchronously. As a consequence, an asynchronous implementation of Algorithm 2 is also guaranteed to converge.

**Remark 1** (Tightness of conditions (5a) and (5b)). *Condition (5a) is a mild requirement. It merely asks for a minimum proportion of agents to move from their current unfavourable action to a better one. Equation (5b), on the other hand, requires the switching to happen sufficiently slowly. Without this condition, the algorithm may not converge, as shown with the following example. Consider  $n = 2$ ,  $u_1(x_1) = 1 - x_1$ ,  $u_2(x_2) = 1 - x_2$ ,  $c_{12} = c_{21} = 0.5$ , and note that  $c_{\min}/L = 0.5$ . Take  $\delta > 0$  small enough and initial condition  $x_1(0) = 0.75 + \delta/2$ ,  $x_2(0) = 0.25 - \delta/2$ . Since  $u_1(0) = 0.25 - \delta/2$  and  $u_2(0) = 0.75 + \delta/2$ , then  $x(0)$  is not an inertial equilibrium. Assume that, as a consequence,  $0.5 + \delta > c_{\min}/L$  units of mass move from action 1 to action 2, resulting in  $x_1(1) = 0.25 - \delta/2$ ,  $x_2(1) = 0.75 + \delta/2$ , and thus  $u_1(1) = 0.75 + \delta/2$ ,  $u_2(1) = 0.25 - \delta/2$ , so  $x(1)$  is not an inertial equilibrium either. A repeated transfer of  $0.5 + \delta$  mass from the action which is worse-off to the one which is better-off results in  $x(2k) = x(0)$  and  $x(2k+1) = x(1)$ . Thus, a slight violation of (5b) brakes the convergence of Algorithm 2.*

## V. EXTENSIONS

We present three modifications of the inertial equilibrium problem, and highlight how the results can be adapted.

**Non-engaging agents.** With the current Definition 2 all the agents are forced to engage, i.e. to choose one of the actions in  $\{1, \dots, n\}$ . Let us now consider an extra action labeled  $e$ , so that the extended actions set is  $\{1, \dots, n, e\}$ . Set  $c_{je} = c_{ej} = 0$  for all  $j \in \{1, \dots, n\}$  and  $u_e(x)$  as some constant value representing, for instance, the utility perceived

when not participating in the game. Within this setup, an agent that does not engage in the game at time  $k = 0$ , will revise his decision at every time-step  $k \geq 1$ , and will rejoin whenever more favourable actions appear. For example, in the ride-hailing application presented in Section VI, action  $e$  could represent electing to temporarily not work as a driver.

**Atomic agents with discrete action set.** Instead of a continuum of agents, one could consider a finite number  $M$  of atomic agents. Each agent possesses unitary mass and can choose only one of the actions  $\{1, \dots, n\}$ . The utility  $u_j$  is then a function of how agents distribute themselves over the actions. The definition of inertial equilibrium requires that no agent  $i \in \{1, \dots, M\}$  has an incentive to switch action, considering the utilities of the alternative actions and the corresponding switching costs. The model with a continuum of agents studied above represents, in a sense, the limiting case obtained as the number of agents  $M$  grows. Since the action space is discrete, the reformulation as a VI is not possible. Nonetheless, one can formulate Algorithm 2 by letting an agent  $i$  switch to an arbitrary action whenever such action is attractive. Convergence is guaranteed upon substituting the expression  $\sum_{j \in \mathcal{E}_i^{\text{out}}(k)} x_{i \rightarrow j}(k)$  in (5a) and (5b) with the number of agents that move at the same time.

**Multi-class inertial equilibrium.** The concept of inertial equilibrium relies on the idea that each agent perceives the same utility  $u_j$  and the same switching costs  $c_{ij}$ . This assumption can be relaxed by introducing different agents' classes. Let  $A$  be the total number of classes, and  $x_i^\alpha$  be the mass of agents belonging to class  $\alpha \in A$  which choose action  $j$ . We denote  $x_i = \sum_{\alpha=1}^A x_i^\alpha$  and  $x^\alpha = \{x_i^\alpha\}_{i=1}^n$ .

**Definition 5.** Consider utilities  $u_i^\alpha : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}$ , switching costs  $c_{ij}^\alpha \geq 0$  and masses  $\gamma^\alpha > 0$ , with  $i, j \in \{1, \dots, n\}$ ,  $\alpha \in \{1, \dots, A\}$ . The vector  $\bar{x} = [\bar{x}^1, \dots, \bar{x}^A] \in \mathbb{R}^{nA}$  is a multi-class inertial equilibrium if  $\bar{x} \geq 0_{nA}$ ,  $\mathbb{1}_n^\top \bar{x}^\alpha = \gamma^\alpha$  for all  $\alpha$ , and

$$\bar{x}_i^\alpha > 0 \implies u_i^\alpha(\bar{x}_r) \geq u_j^\alpha(\bar{x}_r) - c_{ij}^\alpha, \quad \forall j \in \{1, \dots, n\},$$

for all  $i \in \{1, \dots, n\}$  and  $\alpha \in \{1, \dots, A\}$ , where the vector  $\bar{x}_r := \sum_{\alpha=1}^A \bar{x}^\alpha$ .

Note that even though different classes might perceive different utilities at the same action  $i$ , each of these utilities is a function of the sole distribution of the agents on the actions i.e. of the reduced variable  $x_r$ . This is indeed what couples the different classes together. Upon redefining  $\mathcal{S} = \tilde{\mathcal{S}}^1 \times \dots \times \tilde{\mathcal{S}}^A \subset \mathbb{R}^{nA}$  as the Cartesian product of the weighted simplexes  $\tilde{\mathcal{S}}^\alpha = \{x^\alpha \in \mathbb{R}_{\geq 0}^n, \mathbb{1}_n^\top x^\alpha = \gamma^\alpha\}$ , one can redefine  $F : \mathcal{S} \rightarrow \mathbb{R}_{\geq 0}^{nA}$ , where

$$F(x) = [[F_j^\alpha(x)]_{\alpha=1}^A]_{j=1}^n,$$

$$F_j^\alpha(x) = \max_{h \in \{1, \dots, n\}} \left( u_h^\alpha \left( \sum_{\alpha=1}^A x^\alpha \right) - u_j^\alpha \left( \sum_{\alpha=1}^A x^\alpha \right) - c_{jh}^\alpha \right).$$

Using a straightforward extension of the proof of Theorem 1, one can show that the set of multi-class inertial equilibria coincides with the solution set of VI( $\mathcal{S}, F$ ). Theorem 2 about lack of monotonicity also extends to the multi-class case.

Finally, Algorithm 2 can also be modified appropriately to account for the presence of multiple classes, and a similar convergence result to that of Theorem 3 follows.

## VI. APPLICATION: AREA COVERAGE FOR TAXI DRIVERS

In this section we apply the theory developed to the problem of area coverage for taxi drivers. Understanding the spatial behavior of taxi drivers has attracted the interest of the transportation community [22], [23], as it allows to infer information for diverse scopes, including land-use classification [24] and analysis of collective behaviour of a city's population [25].

We focus on the urban area of Hong Kong, as the work [26] provides relevant data for our model. The authors of [26] divide the region of interest into  $n = 18$  neighborhoods, which represent the resources in our game. We assume that a taxi driver in neighborhood  $i$  enjoys the utility  $u_i(x_i)$ , depending on the fraction  $x_i$  of taxi drivers covering the same neighborhood. We aim at determining an equilibrium distribution of the drivers, across the different neighborhoods of the urban area. The problem can be described through the introduction of an undirected graph, where the nodes represent the neighborhoods. We construct an edge from  $i$  to  $j$  (and from  $j$  to  $i$ ) if and only if the two neighborhoods are adjacent. The cost  $c_{ij}$  is taken as the fuel cost of a trip from  $i$  to  $j$  according to [27] and  $c_{ij} = c_{ji}$ , with the fuel cost set as extremely high for non-adjacent neighborhoods, so that movement cannot occur between those. A taxi driver stationing in node  $i$  experiences a utility  $u_i(x_i)$  describing his revenue minus the costs. This takes the form of

$$u_i(x_i) = \alpha_i v_i(x_i) - (1 - v_i(x_i))\beta$$

where  $\alpha_i$  is the average profit per trip starting from location  $i$  (ranging from 30 to 140 HK\$ according to [26]),  $\beta = 6.34$  HK\$ is the operational cost of vacant taxi trips inclusive of fuel costs, rental costs and toll charges associated with the trips. The function  $v_i(x_i)$  describes the percentage of the time a taxi is occupied and according to [28, Eq. (1)] is modeled by

$$v_i(x_i) = 1 - \left( \frac{x_i}{1 + x_i} \right)^{p_i},$$

where  $p_i > 1$  is the number of passengers requesting a taxi at node  $i$ . We select  $p_i$  to be proportional to the values in Figure 3 of [26]. Note that  $v_i(0) = 1$  and  $\lim_{x_i \rightarrow \infty} v_i(x_i) = 0$ , as one would expect. Moreover, through simple algebraic manipulations,  $u_i(x_i)$  can be shown to be decreasing for  $x_i \geq 0$  and to have Lipschitz constant  $L_i = 4(\alpha_i - \beta)p_i \frac{(p_i-1)^{p_i-1}}{(p_i+1)^{p_i+1}}$ . In our numerical study we compare the projection algorithm (Algorithm 1) with the better-response algorithm proposed in Algorithm 2, with stopping criterion  $\|x(k+1) - x(k)\| \leq 10^{-6}$ , and equal neighbour redistribution function  $x_{i \rightarrow j}(k) = \tau x_i(k)$ ,  $j \in \mathcal{E}_i^{\text{out}}(k)$ . For Algorithm 1,  $\rho$  is chosen slightly smaller than  $1/L = 1.7 \cdot 10^{-3}$  as required to achieve convergence by Proposition 2. Similarly,  $\tau$  is chosen slightly smaller than  $c_{\min}/(L\gamma) = 1.4 \cdot 10^{-4}$  in accordance to the requirement of Theorem 3. Table I (top) shows a comparison in terms of iterations needed to reach convergence by both algorithms. Note that a single iteration of Algorithm 1 is more costly

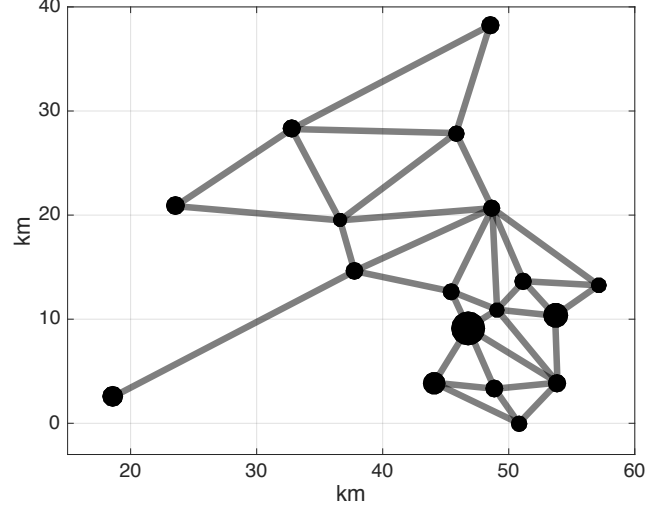


Fig. 2: The equilibrium  $\bar{x}$  achieved by Algorithm 2 with initial condition  $x(0) = \mathbb{1}_n/\gamma$ . The radius of each node is proportional to its utility  $u_i(\bar{x}_i)$ , while the thickness of edge  $(i, j)$  is proportional to the corresponding cost  $c_{ij}$ .

than one of Algorithm 2. Indeed, Algorithm 1 requires the computation of a projection step, while Algorithm 2 requires simple addition and multiplication operations.

We note that the number of required iterations to reach convergence is rather high, due to the small values of  $\rho$  and  $\tau$  imposed by the theoretical bounds of Proposition 2 and Theorem 3. For this reason we perform another simulation with the values  $\rho = \tau = 10^{-2}$ , which provide no theoretical guarantees of convergence. Nonetheless, both algorithms converge in 100 different repetitions with random initial conditions. The number of iterations is reported in Table I (bottom) and is considerably smaller than those in Table I (top). Moreover, Algorithm 1 outperforms Algorithm 2 in the first case, but the viceversa happens in the second case. Finally, Figure 2 shows the steady state distribution of taxi drivers across the  $n$  neighbourhood of Hong Kong, with initial condition  $x(0) = \mathbb{1}_n/\gamma$ .

Algorithm	# iterations mean	# iterations St. Dev.
Alg. 1, $\rho = 10^{-3}$	26 311	3329
Alg. 2, $\tau = 10^{-4}$	152 856	9130
Alg. 1, $\rho = 10^{-2}$	12 672	3472
Alg. 2, $\tau = 10^{-2}$	2168	125

TABLE I: Number of iterations needed to reach convergence with  $\rho = 1.5 \cdot 10^{-3}$ ,  $\tau = 10^{-4}$  (top), and  $\tau = \rho = 10^{-2}$  (bottom). We report mean and standard deviation for 100 repetitions of the two algorithms, starting from random initial conditions in the simplex.

## VII. CONCLUSIONS

We proposed the novel notion of inertial Nash equilibrium to model the cost incurred by agents when switching to an alternative action. While the set of inertial Nash equilibria can be characterized by means of a suitable variational inequality, the resulting operator is often non monotone. Thus, we proposed

a natural dynamics that is distributed, and provably converges to an inertial Nash equilibrium. As future research direction, it would be interesting to provide convergence rate guarantees for Algorithm 2, and more broadly to extend the notion of inertial equilibrium beyond the framework of population games.

## APPENDIX A

### PRELIMINARIES ON VARIATIONAL INEQUALITIES

In the following we present those result on the theory of variational inequality that are used to characterize the equilibrium concepts introduced in Section II.

**Proposition 4** ([21, Prop. 2.3.3]). *Let  $\mathcal{X} \subset \mathbb{R}^n$  be a compact, convex set and  $F : \mathcal{X} \rightarrow \mathbb{R}^n$  be continuous. Then  $\text{VI}(\mathcal{X}, F)$  admits at least one solution.*

The next proposition introduces the KKT system of a variational inequality, which is analogous to the KKT system of an optimization program.

**Proposition 5** ([21, Prop. 1.3.4]). *Assume that the set  $\mathcal{X}$  can be described as  $\mathcal{X} = \{x \in \mathbb{R}^n \mid g(x) \leq \mathbb{0}_m, h(x) = \mathbb{0}_p\}$ , and that it satisfies Slater's constraint qualification in [29, eq. (5.27)]. Then  $\bar{x}$  solves  $\text{VI}(\mathcal{X}, F)$  if and only if there exist  $\bar{\lambda}$  and  $\bar{\mu}$  such that  $(\bar{x}, \bar{\lambda}, \bar{\mu})$  solves the KKT system (6)*

$$F(x) + \nabla_x g(x)\lambda + \nabla_x h(x)\mu = \mathbb{0}_n \quad (6a)$$

$$\mathbb{0}_m \leq \lambda \perp g(x) \leq \mathbb{0}_m \quad (6b)$$

$$h(x) = \mathbb{0}_p. \quad (6c)$$

We next recall the notion of monotonicity, which is a sufficient condition for convergence of a plethora of VI algorithms, see [21, Chapter 12].

**Definition 6** (Monotonicity). *An operator  $F : \mathcal{X} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$  is monotone if for all  $x, y \in \mathcal{X}$ .*

$$(F(x) - F(y))^\top (x - y) \geq 0,$$

**Proposition 6.** [30, Prop. 2.1] *Let  $\mathcal{X} \subseteq \mathbb{R}^n$  be convex. An operator  $F$  is monotone in  $\mathcal{X}$  if and only if for every  $x \in \mathcal{X}$  each generalized Jacobian  $\phi \in \partial F(x)$  is positive semi-definite.*

The definition of generalized Jacobian  $\partial F(x)$  can be found in [31, Definition 2.6.1]; we do not report it here because for our scope it suffices to know that if  $F$  is differentiable in  $x$ , then the generalized Jacobian coincides with the Jacobian, i.e.,  $\partial F(x) = \{\nabla_x F(x)\}$ , with positive-definite interpreted as  $(\nabla_x F(x) + \nabla_x F(x)^\top)/2 \succeq 0$ . We conclude this section with a result on the convexity of the VI solution set.

**Proposition 7** ([21, Thm. 2.3.5]). *Let  $\mathcal{X} \subseteq \mathbb{R}^n$  be closed, convex and  $F : \mathcal{X} \rightarrow \mathbb{R}^n$  be continuous and monotone. Then the solution set of  $\text{VI}(\mathcal{X}, F)$  is convex.*

## APPENDIX B

### PROOFS

#### Proof of Theorem 1

*Proof:* The proof consists in showing that the KKT system of  $\text{VI}(\mathcal{S}, F)$  is equivalent to Definition 2 of inertial

Nash. Since the set  $\mathcal{S}$  satisfies Slater's constraint qualification, by Proposition 5  $\text{VI}(\mathcal{S}, F)$  is equivalent to its KKT system

$$F(x) + \mu \mathbb{1}_n - \lambda = \mathbb{0}_n$$

$$\mathbb{0}_m \leq \lambda \perp x \geq \mathbb{0}_m$$

$$\mathbb{1}_n^\top x = 1$$

where  $\mu \in \mathbb{R}$  is the dual variable corresponding to the constraint  $\mathbb{1}_n^\top x = 1$  and  $\lambda \in \mathbb{R}^n$  is the dual variable corresponding to the constraint  $x \geq \mathbb{0}_n$ . The system (7) can be compactly rewritten as

$$\mathbb{0}_n \leq \mu \mathbb{1}_n + F(x) \perp x \geq \mathbb{0}_n, \quad (8a)$$

$$\mathbb{1}_n^\top x = 1. \quad (8b)$$

Observe that for any  $x \in \mathcal{S}$  there exists  $i^* \in \{1, \dots, n\}$  such that  $F_{i^*}(x) = 0$ . Indeed, setting

$$i^* \in \underset{i \in \{1, \dots, n\}}{\operatorname{argmax}} u_i(x),$$

gives  $F_{i^*}(x) = 0$  by the definition of  $F$  in (4).

It follows that  $\mu < 0$  is not possible, otherwise the non-negativity condition on  $\mu \mathbb{1}_n + F(x)$  is violated. Moreover, since  $F(x) \geq \mathbb{0}_n$ ,  $\mu > 0$  is not possible, as by (8a) this would imply  $x = \mathbb{0}_n$  thus violating (8b). We can conclude that  $\mu = 0$  and (8) becomes

$$\mathbb{0}_n \leq F(x) \perp x \geq \mathbb{0}_n, \quad (9a)$$

$$\mathbb{1}_n^\top x = 1.$$

The system (9) is equivalent to

$$x \in \mathcal{S}, \text{ and}$$

$$x_i > 0 \Rightarrow_{(9a)} u_i(x) \geq u_j(x) - c_{ij}, \forall i, j \in \{1, \dots, n\}.$$

which coincides with Definition 2.

Existence of an inertial equilibrium follows readily from Proposition 4 on the existence of VI solutions. The continuity of the VI operator therein required is satisfied because  $F$  is the point-wise maximum of continuous functions.  $\square$

#### Proof of Theorem 2

*Proof:* The proof is composed of four parts.

1) We first show that there exists  $\hat{x} \in \mathcal{S}_+$  such that  $\hat{x}$  is not an inertial equilibrium (by assumption  $\hat{x}$  belongs to  $\mathcal{S}$  and not necessarily to  $\mathcal{S}_+$ ).

For the sake of contradiction, assume that each  $x \in \mathcal{S}_+$  is an inertial equilibrium. Since  $\hat{x}$  belongs to the closure of  $\mathcal{S}_+$ , we can construct a sequence  $(x(m))_{m=1}^\infty \in \mathcal{S}_+$  such that  $\lim_{m \rightarrow \infty} x(m) = \hat{x}$ . Since each  $x(m)$  is an inertial equilibrium and it is positive, then for all  $i, j$  it holds  $u_i(x(m)) \geq u_j(x(m)) - c_{ij}$ . Taking the limit and exploiting continuity of  $\{u_i\}_{i=1}^n$  we obtain

$$\begin{aligned} \lim_{m \rightarrow \infty} u_i(x(m)) &\geq \lim_{m \rightarrow \infty} u_j(x(m)) - c_{ij}, \\ \Leftrightarrow u_i(\hat{x}) &\geq u_j(\hat{x}) - c_{ij}, \end{aligned} \quad (10)$$

for all  $j, h \in \{1, \dots, n\}$ , hence  $\hat{x}$  is an inertial equilibrium, against the assumption.



2) After establishing the existence of  $\tilde{x} \in S_+$  which is not an inertial equilibrium, we now show that there exists an open ball  $\mathcal{B}_{\tilde{\varepsilon}}(\tilde{x})$  centered around  $\tilde{x}$  of radius  $\tilde{\varepsilon} > 0$  such that none of the points in  $\mathcal{B}_{\tilde{\varepsilon}}(\tilde{x}) \cap S_+$  is an inertial equilibrium. Let us reason again for the sake of contradiction. If for each  $\varepsilon > 0$  there exists an inertial equilibrium in  $\mathcal{B}_{\varepsilon}(\tilde{x}) \cap S_+$ , then we can construct a sequence of inertial equilibria converging to  $\tilde{x}$ . With the same continuity argument used in (10), we can conclude that  $\tilde{x}$  is an inertial equilibrium, which is false by assumption. This demonstrates the existence of  $\tilde{\varepsilon} > 0$  such that none of the points in  $\mathcal{B}_{\tilde{\varepsilon}}(\tilde{x}) \cap S_+$  is an inertial equilibrium. By Rademacher's theorem [32, Thm. 2.14], Lipschitzianity of  $\{u_i\}_{i=1}^n$  guarantees<sup>3</sup> existence of  $x^* \in \mathcal{B}_{\tilde{\varepsilon}}(\tilde{x}) \cap S_+$  such that  $F$  is differentiable at  $x^*$ .

3) The previous part guarantees differentiability of  $F$  at a point  $x^* \in S_+$  which is not an inertial equilibrium. This third part is dedicated to showing that there exist  $i^*, j^* \in \{1, \dots, n\}$  such that  $i^* \in \mathcal{A}(j^*, x^*)$  and  $\mathcal{A}(i^*, x^*) = \{i^*\}$ , where we denote

$$\mathcal{A}(k, x) := \operatorname{argmax}_{\ell \in \{1, \dots, n\}} \{u_{\ell}(x) - u_k(x) - c_{k\ell}\}.$$

Since  $x^*$  is not an inertial equilibrium, then there exist  $\ell_1, \ell_2$  such that

$$u_{\ell_1}(x^*) < u_{\ell_2}(x^*) - c_{\ell_1\ell_2}. \quad (11)$$

Condition (11) is equivalent to  $\ell_2 \in \mathcal{A}(\ell_1, x^*)$  and  $\ell_1 \notin \mathcal{A}(\ell_1, x^*)$ . If  $\mathcal{A}(\ell_2, x^*) = \{\ell_2\}$  then the statement is proven with  $j^* = \ell_1, i^* = \ell_2$ , otherwise there exists  $\ell_3 \in \mathcal{A}(\ell_2, x^*) \setminus \{\ell_2\}$ . Note that it cannot be  $\ell_3 = \ell_1$ , because this means

$$u_{\ell_2}(x^*) \leq u_{\ell_1}(x^*) - c_{\ell_2\ell_1},$$

which together with (11) results in

$$u_{\ell_1}(x^*) < u_{\ell_1}(x^*) - c_{\ell_2\ell_1} - c_{\ell_1\ell_2},$$

which is not possible, because  $c_{\ell_1\ell_2}, c_{\ell_2\ell_1} \geq 0$  by assumption. Hence we established that  $\ell_3 \neq \ell_1$ . If  $\mathcal{A}(\ell_3, x^*) = \{\ell_3\}$  then the statement is proven with  $j^* = \ell_2, i^* = \ell_3$ , otherwise there exists  $\ell_4 \notin \{\ell_1, \ell_2, \ell_3\}$  such that  $\ell_4 \in \mathcal{A}(\ell_3, x^*)$ . Since there are only  $n$  different actions, by continuing the chain of reasoning we conclude that there exists  $k \in \{2, \dots, n\}$  such that  $\ell_k \in \mathcal{A}(\ell_{k-1}, x^*)$  and  $\mathcal{A}(\ell_k, x^*) = \{\ell_k\}$ , thus proving the statement with  $j^* = \ell_{k-1}$  and  $i^* = \ell_k$ .

We now proceed to show that not only  $i^* \in \mathcal{A}(j^*, x^*)$ , but actually  $\mathcal{A}(j^*, x^*) = \{i^*\}$ . For the sake of contradiction, assume that there exists  $\ell \neq i^*$  such that  $\ell \in \mathcal{A}(j^*, x^*)$ . This means that  $F_{j^*}(x^*) = u_{i^*}(x^*) - u_{j^*}(x^*) - c_{j^*i^*} = u_{\ell}(x^*) - u_{j^*}(x^*) - c_{j^*\ell}$ . Then consider the vector of the

canonical basis  $e_{i^*} \in \mathbb{R}^n$  and compute

$$\begin{aligned} \lim_{t \rightarrow 0^+} \frac{F_{j^*}(x^* + te_{i^*}) - F_{j^*}(x^*)}{t} &= \\ \lim_{t \rightarrow 0^+} \frac{[u_{\ell}(x^*) - u_{j^*}(x^*) - c_{j^*\ell}] - [u_{i^*}(x^*) - u_{j^*}(x^*) - c_{j^*i^*}]}{t} &= 0, \end{aligned} \quad (12)$$

where the first equality holds because for  $t > 0$  we have

$$\begin{aligned} u_{i^*}(x^* + te_{i^*}) - u_{j^*}(x^*) - c_{j^*i^*} &< u_{i^*}(x^*) - u_{j^*}(x^*) - c_{j^*i^*} \\ &= u_{\ell}(x^*) - u_{j^*}(x^*) - c_{j^*\ell}, \end{aligned}$$

due to  $\nabla_{x_{i^*}} u_{i^*}(x^*) < 0$  by assumption. Moreover,

$$\begin{aligned} \lim_{t \rightarrow 0^-} \frac{F_{j^*}(x^* + te_{i^*}) - F_{j^*}(x^*)}{t} &= \\ \lim_{t \rightarrow 0^-} \frac{[u_{i^*}(x^* + te_{i^*}) - u_{j^*}(x^*) - c_{j^*i^*}] - [u_{i^*}(x^*) - u_{j^*}(x^*) - c_{j^*i^*}]}{t} &= \\ = \lim_{t \rightarrow 0^-} \frac{u_{i^*}(x^* + te_{i^*}) - u_{i^*}(x^*)}{t} &= \nabla_{x_{i^*}} u_{i^*}(x^*) < 0, \end{aligned} \quad (13)$$

where the first equality holds because for  $t < 0$  we have

$$\begin{aligned} u_{i^*}(x^* + te_{i^*}) - u_{j^*}(x^*) - c_{j^*i^*} &> u_{i^*}(x^*) - u_{j^*}(x^*) - c_{j^*i^*} \\ &= u_{\ell}(x^*) - u_{j^*}(x^*) - c_{j^*\ell}, \end{aligned}$$

due to  $\nabla_{x_{i^*}} u_{i^*}(x^*) < 0$  by assumption. From (12) and (13) we obtain that  $F_{j^*}$  is not differentiable at  $x^*$ , against what proved in the second part. Hence we must conclude that there cannot exist  $\ell \neq i^*$  such that  $\ell \in \mathcal{A}(j^*, x^*)$ , thus  $\mathcal{A}(j^*, x^*) = \{i^*\}$ .

4) Since  $F$  is differentiable in  $x^*$  by the second part of the proof, then  $\partial F(x^*) = \{\nabla_x F(x^*)\}$  is a singleton. As  $\mathcal{A}(j^*, x^*) = \mathcal{A}(i^*, x^*) = \{i^*\}$  by the third part of the proof, then

$$\begin{aligned} u_{i^*}(x^*) - c_{j^*i^*} &> u_{\ell}(x^*) - c_{j^*\ell}, \quad \forall \ell \neq i^*, \\ u_{i^*}(x^*) - c_{i^*i^*} &> u_{\ell}(x^*) - c_{i^*\ell}, \quad \forall \ell \neq i^*. \end{aligned} \quad (14)$$

As a consequence of (14) there exists a small enough open ball around  $x^*$  where  $F_{i^*}(x^*) = u_{i^*}(x^*) - u_{i^*}(x^*) - c_{i^*i^*} = 0$  and  $F_{j^*}(x^*) = u_{i^*}(x^*) - u_{j^*}(x^*) - c_{j^*i^*}$ . Thus

$$\begin{aligned} [\nabla_x F(x^*)]_{i^*j^* \times i^*j^*} &= \\ \begin{bmatrix} \frac{\partial F_{i^*}(x^*)}{\partial x_{i^*}} & \frac{\partial F_{i^*}(x^*)}{\partial x_{j^*}} \\ \frac{\partial F_{j^*}(x^*)}{\partial x_{i^*}} & \frac{\partial F_{j^*}(x^*)}{\partial x_{j^*}} \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ \nabla_{x_{i^*}} u_{i^*}(x^*) & -\nabla_{x_{j^*}} u_{j^*}(x^*) \end{bmatrix}, \end{aligned}$$

whose symmetric part has determinant  $0 \cdot \nabla_{x_{j^*}} u_{j^*}(x^*) - (\nabla_{x_{i^*}} u_{i^*}(x^*))^2/4 < 0$ , which makes  $[\nabla_x F(x^*)]_{i^*j^* \times i^*j^*}$  indefinite. Thus  $\nabla_x F(x^*)$  itself is indefinite and  $F$  is not monotone in  $S$  due to Proposition 6.  $\square$

## Proof of Proposition 2

*Proof:* Algorithm 1 is the projection algorithm in [21, Alg. 12.1.1], applied to  $\text{VI}(S, -u)$ . A solution of  $\text{VI}(S, -u)$  exists by Proposition 4. The operator  $-u$  is monotone in  $S$ , because  $\theta$  is concave [33, eq. (12)]. Moreover, due to existence of  $\theta$ ,  $L$ -Lipschitzianity is equivalent to  $(1/L)$ -cocoercitivity [34, Thm. 18.15]. Then, for  $\rho < 2/L$ , Algorithm 1 is guaranteed to

<sup>3</sup>Rademacher's theorem assumes  $F$  to be defined on an open subset of  $\mathbb{R}^n$ , but  $S_+$  is not open in  $\mathbb{R}^n$ . Indeed, one just needs to define  $F$  on the  $n-1$  dimensional open set  $\{x \in \mathbb{R}_{>0}^{n-1} | \mathbf{1}_{n-1}^\top x < 1\}$ , by using  $x_n = 1 - \sum_{j=1}^{n-1} x_j$  and then apply the Rademacher's Theorem to conclude existence of a differentiable point in  $\{x \in \mathbb{R}_{>0}^{n-1} | \mathbf{1}_{n-1}^\top x < 1\}$  which implies existence of a differentiable point in the original  $S_+$ .

converge to a solution of  $\text{VI}(\mathcal{S}, -u)$  by [21, Thm. 12.1.8]. The final claim follows by observing that any Wardrop equilibrium is also an inertial Wardrop equilibrium (Lemma 1).  $\square$

### Proof of Theorem 3

*Proof:* First, observe that if  $x(0) \in \mathcal{S}$ , then  $x(k)$  remains in  $\mathcal{S}$  for all  $k \geq 1$ . This is consequence of the two following observations. i) At every fixed time-step  $k$ , and for every pair  $i, j$  with  $j \in \mathcal{E}_i^{\text{out}}(k)$ , the mass  $x_{i \rightarrow j}(k)$  is removed from node  $i$  and simultaneously added to node  $j$  (see Algorithm 2). Therefore, the total mass must be conserved at each iteration, and so it must be  $\sum_{i \in \{1, \dots, n\}} x_i(k) = \sum_{i \in \{1, \dots, n\}} x_i(0) = 1$ . ii) For every node  $i \in \{1, \dots, n\}$ , the evolution of  $x_i(k)$ , as dictated by Algorithm 2, can be compactly written as

$$x_i(k+1) = x_i(k) - \sum_{j \in \mathcal{E}_i^{\text{out}}(k)} x_{i \rightarrow j}(k) + \sum_{\ell \text{ s.t. } i \in \mathcal{E}_\ell^{\text{out}}(k)} x_{\ell \rightarrow i}(k).$$

Since by assumption  $\sum_{j \in \mathcal{E}_i^{\text{out}}(k)} x_{i \rightarrow j}(k) \leq x_i(k)$  for every time-step  $k$ , we have that  $x_i(k+1) \geq \sum_{\ell \text{ s.t. } i \in \mathcal{E}_\ell^{\text{out}}(k)} x_{\ell \rightarrow i}(k) \geq 0$ , where the last inequality follows from  $x_{\ell \rightarrow i}(k) \geq 0$ . Repeating the reasoning for every  $k$  ensures that  $x_i(k) \geq 0$  at every time-step. Finally, since  $\sum_{j \in \mathcal{E}_i^{\text{out}}(k)} x_{i \rightarrow j}(k) \leq x_i(k)$ , it must be that  $x_{\ell \rightarrow i}(k) \leq x_\ell(k)$ . Therefore  $\sum_{\ell \text{ s.t. } i \in \mathcal{E}_\ell^{\text{out}}(k)} x_{\ell \rightarrow i}(k) \leq \sum_{\ell \text{ s.t. } i \in \mathcal{E}_\ell^{\text{out}}(k)} x_\ell(k) \leq \sum_{\ell \neq i} x_\ell(k)$ . Hence  $x_i(k+1) \leq \sum_{\ell \in \{1, \dots, n\}} x_\ell(k) - \sum_{j \in \mathcal{E}_i^{\text{out}}(k)} x_{i \rightarrow j}(k) \leq 1$ , where the last inequality follows from the fact that  $\sum_{\ell \in \{1, \dots, n\}} x_\ell(k) = 1$  (as shown above) and from the fact that  $x_{i \rightarrow j}(k) \geq 0$ .

We now move our attention to proving the desired convergence statement. To do so, we will show that  $x(k) \rightarrow \bar{x}$  such that  $\mathcal{E}_i^{\text{out}}(\bar{x}) = \emptyset$  for all  $i \in \{1, \dots, n\}$ , thanks to the equivalence in Proposition 3. Let us denote for brevity  $u_i(k) := u_i(x_i(k))$  and define  $\mu(k) = \min_{i \in \{1, \dots, n\}} u_i(k)$ . We show in the following that  $\mu(k)$  is a non-decreasing sequence.

First, for any action  $i$  we have  $x_i(k+1) - x_i(k) \leq c_{\min}/L - \varepsilon$  due to (5b). Then we can bound the maximum utility decrease

$$\begin{aligned} u_i(k+1) - u_i(k) &\geq -L|x_i(k+1) - x_i(k)| \\ &\geq -L(c_{\min}/L - \varepsilon) = -c_{\min} + L\varepsilon =: -\beta c_{\min}, \end{aligned} \quad (15)$$

where the first inequality follows by Lipschitz continuity and we define  $\beta := 1 - (L\varepsilon)/c_{\min} \in ]0, 1[$ .

Secondly, note that if some action  $i$  faces a utility decrease, that is, if  $u_i(k+1) < u_i(k)$ , then it must be  $x_i(k+1) > x_i(k)$ , because  $u_i$  is non-increasing. Then there exists  $j$  such that  $i \in \mathcal{E}_j^{\text{out}}(x(k))$ . It follows that

$$\begin{aligned} i \text{ faces utility decrease at step } k &\Rightarrow \\ u_i(k) &> u_j(k) + c_{ji} \geq \mu(k) + c_{\min}. \end{aligned} \quad (16)$$

Combining (15) with (16) we obtain

$$\begin{aligned} i \text{ faces utility decrease at step } k &\Rightarrow \\ u_i(k+1) &> \mu(k) + (1 - \beta)c_{\min}, \end{aligned}$$

which implies  $\mu(k+1) \geq \mu(k)$ . Since  $\mu(k)$  is non-decreasing and bounded ( $\{u_i\}_{i=1}^n$  are continuous functions in a compact

set), there exists a value  $\mu^*$  such that

$$\lim_{k \rightarrow \infty} \mu(k) = \mu^*. \quad (17)$$

We show in the following that there exists an action  $i^*$  such that

$$\lim_{k \rightarrow \infty} u_{i^*}(k) = \mu^*. \quad (18)$$

As  $\lim_{k \rightarrow \infty} \mu(k) = \mu^*$ , there exists  $\hat{k}$  such that

$$\mu(k) > \mu^* - c_{\min}(1 - \beta)/2, \quad \forall k \geq \hat{k}. \quad (19)$$

Then

$$\begin{aligned} i \text{ faces utility decrease at step } k \geq \hat{k} &\Rightarrow \\ u_i(k) &\geq \mu^* - c_{\min}(1 - \beta)/2 + c_{\min} \\ &= \mu^* + c_{\min}(1 + \beta)/2, \end{aligned} \quad (20)$$

where the first inequality follows from combining (16) and (19). Combining (15) and (20) we obtain

$$\begin{aligned} i \text{ faces utility decrease at step } k \geq \hat{k} &\Rightarrow \\ u_i(k+1) &\geq \mu^* - c_{\min}(1 - \beta)/2 + c_{\min}(1 - \beta) \\ &= \mu^* + c_{\min}(1 - \beta)/2. \end{aligned} \quad (21)$$

Figure 3 illustrates inequalities (20) and (21).

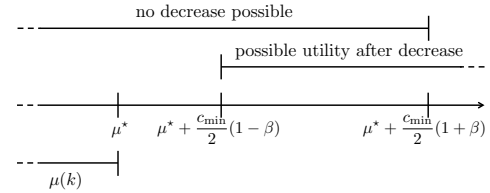


Fig. 3: Illustration of  $\mu(k) \rightarrow \mu^*$  from below and of inequalities (20) and (21) after iteration  $\hat{k}$  (with  $\beta = 0.5$ ).

Combining inequalities (20) and (21) we obtain that

$$\begin{aligned} \exists k_1 \geq \hat{k} \text{ such that } u_i(k_1) &\geq \mu^* + \rho > \mu^* \Rightarrow \\ u_i(k) &\geq \min\{\mu^* + \rho, \mu^* + c_{\min}(1 - \beta)/2\} \text{ for all } k \geq k_1. \end{aligned} \quad (22)$$

It then follows

$$\exists k_1 \geq \hat{k} \text{ such that } u_i(k_1) > \mu^* \Rightarrow \lim_{k \rightarrow \infty} u_i(k) \neq \mu^*. \quad (23)$$

By (23) and (17) it follows that there exists at least an action  $i^*$  such that  $u_{i^*}(k) \leq \mu^*$  for all  $k \geq \hat{k}$ . Using again (17) and the “squeeze theorem” [35, Thm. 3.3.6], we can conclude that  $i^*$  satisfies (18). Upon defining

$$\mathcal{E}_j^{\text{in}}(x) = \{i \in \{1, \dots, n\} \text{ s.t. } j \in \mathcal{E}_i^{\text{out}}(x)\},$$

for any  $j \in \{1, \dots, n\}$  and  $x \in \mathcal{S}$ , we note that the set  $\mathcal{E}_{i^*}^{\text{in}}(x(k))$  is empty for  $k \geq \hat{k}$  due to (16) and  $u_{i^*}(k) \leq \mu^*$ . In words, no other action can envy  $i^*$  after step  $\hat{k}$ . This implies that  $u_{i^*}(k)$  is a non-decreasing sequence, and in turn  $x_{i^*}(k)$  is a non-increasing sequence. As a consequence

$$\lim_{k \rightarrow \infty} x_{i^*}(k) = \bar{x}_{i^*} \geq 0. \quad (24)$$

If  $\bar{x}_{i^*} = 0$ , then clearly  $\mathcal{E}_{i^*}^{\text{out}}(\bar{x}_{i^*}, x_{-i^*}) = \emptyset$  by definition, for any  $x_{-i^*}$ . If instead  $\bar{x}_{i^*} > 0$ , since  $x_{i^*}(k+1) \leq (1 - \tau)x_{i^*}(k)$  due to (5a), then convergence is achieved in a finite number of

steps. In other words, there exists  $\tilde{k}$  such that  $x_{i^*}(k) = \bar{x}_{i^*}$  for all  $k \geq \tilde{k}$ . In this case, for  $k \geq \tilde{k}$  not only  $\mathcal{E}_{i^*}^{\text{in}}(x(k)) = \emptyset$ , but also  $\mathcal{E}_{i^*}^{\text{out}}(x(k)) = \emptyset$ , because otherwise  $i^*$  would encounter a mass decrease.

Having concluded that there exists  $i^* \in \{1, \dots, n\}$  such that its mass converges (in a finite number of steps if  $\bar{x}_{i^*} > 0$ ), we propose a last argument to show that there exists  $j^* \in \{1, \dots, n\} \setminus \{i^*\}$  such that its mass converges to  $\bar{x}_{j^*}$  (in a finite number of steps if  $\bar{x}_{i^*}, \bar{x}_{j^*} > 0$ ). Applying the same argument recursively to  $\{1, \dots, n\} \setminus \{i^*, j^*\}$  concludes the proof.

The last argument distinguishes two cases:  $\bar{x}_{i^*} > 0$  and  $\bar{x}_{i^*} = 0$ . In the first case  $\bar{x}_{i^*} > 0$ , we already showed that there exists  $\tilde{k}$  such that  $\mathcal{E}_{i^*}^{\text{in}}(x(k)) = \mathcal{E}_{i^*}^{\text{out}}(x(k)) = \emptyset$  for all  $k > \tilde{k}$ . Then action  $i^*$  has no interaction with any the other action and considering  $k \geq \tilde{k}$  we apply to  $\{1, \dots, n\} \setminus i^*$  the previous reasoning until equation (24) to show that there is an action  $j^* \in \{1, \dots, n\} \setminus \{i^*\}$  with mass that converges to  $\bar{x}_{j^*}$  (in a finite number of steps if  $\bar{x}_{j^*} > 0$ ).

In the second case  $\bar{x}_{i^*} = 0$ . Even though  $\mathcal{E}_{i^*}^{\text{out}}$  does not become the empty set at any finite iteration  $k$ , the mass  $x_{i^*}$  becomes so small that transferring mass to the other  $n - 1$  actions does not have an influence on their convergence. Proving this requires a cumbersome analysis that does not add much to the intuition already provided. Let us denote  $\eta(k) = \min_{j \in \{1, \dots, n\} \setminus \{i^*\}} u_j(k)$ . Contrary to  $\mu(k)$ , the sequence  $\eta(k)$  is not non-decreasing in general because the analogous of (16) does not hold, as action  $i^*$  could transfer some of its mass to  $\{1, \dots, n\} \setminus \{i^*\}$  thus making their utilities decrease. Nonetheless, we show that there exists  $\eta^*$  such that

$$\lim_{k \rightarrow \infty} \eta(k) = \eta^*. \quad (25)$$

To this end, we fix  $\epsilon > 0$  and we show that there exists  $k^*$  such that  $|\eta(k) - \eta^*| < \epsilon$  for all  $k \geq k^*$ . By definition of  $\lim_{k \rightarrow \infty} x_{i^*}(k) = 0$ , there exists  $k_\infty$  such that

$$x_{i^*}(k) < \epsilon/(2L), \quad \forall k \geq k_\infty. \quad (26)$$

Let us now construct the sequence

$$\begin{aligned} \eta^0(k) &= \eta(k) + \delta(k), \\ \delta(k+1) &= \delta(k) + \max\{0, \eta(k) - \eta(k+1)\}, \quad \delta(k_\infty) = 0. \end{aligned}$$

In words, the sequence  $\delta(k)$  accumulates the (absolute value of the) decreases of  $\eta(k)$  due to  $i^*$ , and summing it to  $\eta(k)$  results in a sequence  $\eta^0(k)$  which is non-decreasing and bounded from above, hence it admits a limit  $\eta^*$ . By definition, there exists  $k^0$  such that  $\eta^0(k) > \eta^* - \epsilon/2$  for all  $k \geq k^0$ . Moreover,  $\delta(k+1) - \delta(k) = \max\{0, \eta(k) - \eta(k+1)\} > 0$  only if  $\mathcal{E}_{i^*}^{\text{out}}(x(k)) \neq \emptyset$  and in this case  $\max\{0, \eta(k) - \eta(k+1)\} \leq L \cdot \sum_{j \neq i^*} x_{i^* \rightarrow j}(k)$ . In words, the only way  $\eta(k)$  can decrease is if action  $i^*$  transfers some mass to the others, and even then we have a bound on the utility decrease that this can cause. Summing up

$$\begin{aligned} \lim_{k \rightarrow \infty} \delta(k) &= \sum_{k=k_\infty}^{\infty} \max\{0, \eta(k) - \eta(k+1)\} \\ &\leq L x_{i^*}(k_\infty) \stackrel{(26)}{\leq} \epsilon/2, \end{aligned}$$

hence, since  $\delta(k)$  is non-decreasing,  $\delta(k) < \epsilon/2$  for all  $k \geq k_\infty$ . Then for  $k \geq \max\{k^0, k_\infty\}$  it holds

$$\begin{aligned} \eta^* - \eta(k) &= \eta^* - \eta^0(k) + \eta^0(k) - \eta(k) \\ &= \underbrace{\eta^* - \eta^0(k)}_{< \epsilon/2} + \underbrace{\eta^0(k) - \eta(k)}_{< \epsilon/2} < \epsilon \end{aligned}$$

which proves (25).

Finally, we want to show that there exists  $j^* \in \{1, \dots, n\} \setminus \{i^*\}$  such that

$$\lim_{k \rightarrow \infty} u_{j^*}(k) = \eta^*. \quad (27)$$

Consider an action  $\ell \neq i^*$  such that

$$\lim_{k \rightarrow \infty} u_\ell(x_\ell(k)) \neq \eta^*. \quad (28)$$

Since  $\eta(k) \rightarrow \eta^*$ , then  $\max\{0, \eta(k) - \eta(k+1)\} \rightarrow 0$  as  $k \rightarrow \infty$ . This, together with  $\eta(k) \rightarrow \eta^*$ , implies that condition (28) is equivalent to the existence of  $\theta > 0$  such that for all  $k' \geq 0$  there exists  $k'' \geq k'$  such that

$$u_\ell(k'') > \eta^* + \theta. \quad (29)$$

There are two possibilities in which  $\ell$  can face a utility decrease after  $k''$ , namely through a mass transfer from some action  $\{1, \dots, n\} \setminus \{i^*, \ell\}$  or through a mass transfer from action  $i^*$ . If the mass transfer happens through some action  $\{1, \dots, n\} \setminus \{i^*, \ell\}$ , we can use the same argument of Figure 3 and in particular of implication (22) to conclude from (29) that

$$u_\ell(k) \geq \min\{\eta^* + \theta, \eta^* + c_{\min}(1 - \beta)/2\}, \quad \forall k \geq k''. \quad (30)$$

If instead the mass transfer happens through  $i^*$ , by  $x_{i^*}(k) \rightarrow 0$  one can take  $k'$  such that

$$x_{i^*}(k) < \theta/(2L), \quad \forall k \geq k' \quad (31)$$

and take  $k''$  such that (29) holds. Then

$$u_\ell(k) \geq u_\ell(k'') - L \frac{\theta}{2L} > \eta^* + \theta - \frac{\theta}{2} = \eta^* + \frac{\theta}{2}. \quad (32)$$

for all  $k \geq k''$ , where the first inequality holds due to Lipschitz continuity and to (31), while the second inequality holds due to (29). We can conclude that if (28) holds for action  $\ell$ , then either (30) or (32) holds. Consequently, after  $k''$  action  $\ell$  does not attain the minimum  $\eta(k)$ . If (28) holds for all  $\ell \in \{1, \dots, n\} \setminus i^*$ , then the minimum  $\eta(k)$  is not attained by any action after  $k''$ , which is a contradiction. Then there must exist  $j^*$  such that (27) holds. With the same argument that led to (24), we can conclude that there exists  $\bar{x}_{j^*} \geq 0$  such that  $\lim_{k \rightarrow \infty} x_{j^*}(k) = \bar{x}_{j^*} \geq 0$ . As done for  $i^*$ , we can conclude that  $\mathcal{E}_{j^*}^{\text{out}} = \emptyset$ .  $\square$

## REFERENCES

- [1] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 1, pp. 67–78, 2013.
- [2] T. Alpcan, T. Başar, R. Srikant, and E. Altman, "CDMA uplink power control as a noncooperative game," *Wireless Networks*, vol. 8, no. 6, pp. 659–670, 2002.
- [3] R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Mathematics of Operations Research*, vol. 29, no. 3, pp. 407–435, 2004.

- [4] J. G. Wardrop, "Some theoretical aspects of road traffic research." *Proceedings of the institution of civil engineers*, vol. 1, no. 3, pp. 325–362, 1952.
- [5] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press Cambridge, 2007.
- [6] F. Dieleman, *Households and housing: Choice and outcomes in the housing market*. Routledge, 2017.
- [7] S. Schulmeister, "A general financial transaction tax motives, revenues, feasibility and effects," *WIFO*, March, 2008.
- [8] L. Coch and J. R. French Jr, "Overcoming resistance to change," *Human relations*, vol. 1, no. 4, pp. 512–532, 1948.
- [9] S. Oreg, "Resistance to change: Developing an individual differences measure." *Journal of applied psychology*, vol. 88, no. 4, p. 680, 2003.
- [10] O. Oyeniyi and A. Abiodun, "Switching cost and customers loyalty in the mobile phone market: The Nigerian experience," *Business Intelligence Journal*, vol. 3, no. 1, pp. 111–121, 2010.
- [11] W. H. Sandholm, "Potential games with continuous player sets," *Journal of Economic Theory*, vol. 97, no. 1, pp. 81–108, 2001.
- [12] A. Nagurney, "Migration equilibrium and variational inequalities," *Economics letters*, vol. 31, no. 1, pp. 109–112, 1989.
- [13] W. H. Sandholm, *Population games and evolutionary dynamics*. The MIT press, 2010.
- [14] J. H. Nachbar, "Evolutionary selection dynamics in games: Convergence and limit properties," *International journal of game theory*, vol. 19, no. 1, pp. 59–89, 1990.
- [15] R. Cressman and Y. Tao, "The replicator equation and other game dynamics," *Proceedings of the National Academy of Sciences*, vol. 111, no. Supplement 3, pp. 10810–10817, 2014.
- [16] L. Zino, G. Como, and F. Fagnani, "On imitation dynamics in potential population games," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 757–762.
- [17] A. Nagurney, J. Pan, and L. Zhao, "Human migration networks," *European journal of operational research*, vol. 59, no. 2, pp. 262–274, 1992.
- [18] —, "Human migration networks with class transformations," in *Structure and Change in the Space Economy*. Springer, 1993, pp. 239–258.
- [19] P. Hartman and G. Stampacchia, "On some non-linear elliptic differential-functional equations," *Acta mathematica*, vol. 115, no. 1, pp. 271–310, 1966.
- [20] R. W. Cottle, "Nonlinear programs with positively bounded Jacobians," *SIAM Journal on Applied Mathematics*, vol. 14, no. 1, pp. 147–158, 1966.
- [21] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.
- [22] P. S. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi GPS traces," in *International Conference on Pervasive Computing*. Springer, 2012, pp. 57–72.
- [23] B. Li, D. Zhang, L. Sun, C. Chen, S. Li, G. Qi, and Q. Yang, "Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 63–68.
- [24] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li, "Land-use classification using taxi GPS traces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 113–123, 2013.
- [25] P. S. Castro, D. Zhang, C. Chen, S. Li, and G. Pan, "From taxi GPS traces to social and community dynamics: A survey," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, p. 17, 2013.
- [26] R. C. P. Wong, W. Y. Szeto, S. Wong, and H. Yang, "Modelling multi-period customer-searching behaviour of taxi drivers," *Transportmetrica B: Transport Dynamics*, vol. 2, no. 1, pp. 40–59, 2014.
- [27] (2017) Viamichelin. [Online]. Available: <https://www.viamichelin.com>
- [28] N. Buchholz, "Spatial equilibrium, search frictions and efficient regulation in the taxi industry," Working paper, Tech. Rep., 2015.
- [29] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [30] S. Dinh The Luc, Schaible, "Generalized monotone nonsmooth maps," *Journal of Convex Analysis*, vol. 3, pp. 195–206, 1996.
- [31] F. H. Clarke, *Optimization and nonsmooth analysis*. SIAM, 1990.
- [32] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*. Clarendon Press Oxford, 2000, vol. 254.
- [33] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 35–49, 2010.
- [34] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2010.
- [35] H. H. Sohrab, *Basic real analysis*. Springer, 2003.



with applications to traffic networks and to smart charging of electric vehicles.



From March to August 2017 he has been a visiting scholar at the University of California, Santa Barbara. Dario's research interests are at the interface between distributed control and game theory. Applications include multiagent systems, smart cities and traffic networks.



**Bolutife Ogunsula** is a software engineer at Bloomberg LP, who completed a master's degree in Robotics, Systems and Control Engineering from ETH Zurich. Prior to that, he got his Bachelor's degree in Electrical and Electronics Engineering from the University of Lagos, and he worked as a software engineer for Codility.



**John Lygeros** completed a B.Eng. degree in electrical engineering in 1990 and an M.Sc. degree in Systems Control in 1991, both at Imperial College of Science Technology and Medicine, London, UK. In 1996 he obtained a Ph.D. degree from the Electrical Engineering and Computer Sciences Department, University of California, Berkeley. During the period 1996-2000 he held a series of research appointments. Between 2000 and 2003 he was a University Lecturer at the Department of Engineering, University of Cambridge, UK. Between 2003 and 2006 he was an

Assistant Professor at the Department of Electrical and Computer Engineering, University of Patras, Greece. In July 2006 he joined the Automatic Control Laboratory at ETH Zürich, first as an Associate Professor, and since January 2010 as a Full Professor. Since 2009 he is serving as the Head of the Automatic Control Laboratory and since 2015 as the Head of the Department of Information Technology and Electrical Engineering. His research interests include modelling, analysis, and control of hierarchical, hybrid, and stochastic systems, with applications to biochemical networks, automated highway systems, air traffic management, power grids and camera networks. John Lygeros is a Fellow of the IEEE, and a member of the IET and the Technical Chamber of Greece.