# The Nash Equilibrium With Inertia in Population Games

Basilio Gentile, Dario Paccagnan, Bolutife Ogunsula, and John Lygeros, *Fellow, IEEE*

***Abstract*—In the traditional game-theoretic set up, where agents select actions and experience corresponding utilities, a Nash equilibrium is a configuration where no agent can improve their utility by unilaterally switching to a different action. In this article, we introduce the novel notion of inertial Nash equilibrium to account for the fact that in many practical situations switching action does 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 switching to a different action, while taking into account the cost of such switch. First, we show that the set of inertial Nash equilibria contains all the Nash equilibria, is in general nonconvex, and can be characterized as a solution to a variational inequality. We then argue that classical algorithms for computing Nash equilibria cannot be used in the presence of switching costs. Finally, we propose a better-response dynamics algorithm and prove its convergence to an inertial Nash equilibrium. We apply our results to study the taxi drivers' distribution in Hong Kong.**

***Index Terms*—distributed algorithms, games, multiagent systems, Nash equilibrium, vehicle routing.**

## I. Introduction

GAME theory has originated to model and describe the interaction of multiple decision makers, or agents. One of the goals is to determine whether decision makers will come to some form of equilibrium, the most common being the Nash equilibrium. A set of strategies constitutes a Nash equilibrium if no agent benefits by unilaterally deviating form 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 concept of Nash equilibrium does not account for the cost incurred by agents when switching to a different action, in practical situations decision makers often incur a physical, psychological, or monetary cost for such switch. This is the case, for example, when relocating to a new neighborhood [6], or when switching financial strategy in the stock market [7]. If the decision makers are humans, the psychological resistance to change has been documented and studied at the professional and organizational level [8], at the individual and private level [9], and 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 switching from action $i$ to action $j$ with the nonnegative coefficients $c_{ij}$. We define an *inertial Nash equilibrium* as a distribution over the action space where no agent has any incentive to unilaterally switch action, when accounting not only for utility gain but also for switching cost.

Utilizing the notion of Nash equilibrium would incorrectly label many distributions as "not of equilibrium," in spite of the fact that agents would stick to them due to the presence of switching costs.

### A. Contributions

The main contributions of this article are as follows.

1) We introduce the notion of inertial Nash equilibrium, which leverages switching costs to realistically describe many applications arising in competitive decision making.
2) We show that the set of inertial Nash equilibria can be characterised through a variational inequality (VI) (see Theorem 1) and prove a strong negative result: the corresponding VI is nonmonotone in all nondegenerate instances with decreasing utility functions (see Theorem 2). Thus, existing variational inequalities algorithms are not suitable for computing an inertial Nash equilibrium.

Manuscript received June 23, 2020; accepted November 18, 2020. Date of publication December 11, 2020; date of current version December 3, 2021. This work was supported in part by the European Research Council (ERC) under Project OCAL under Grant 787845 and in part by the Swiss National Science Foundation (SNSF) under Grant #P2EZP2-181618. Recommended by Associate Editor R. M. Jungers. *(Corresponding author: Dario Paccagnan.)*

Basilio Gentile is with the Circuit Mind, London SE1 0HS, U.K. (e-mail: gentileb@control.ee.ethz.ch).

Dario Paccagnan is with the Department of Computing, Imperial College London, London SW72AZ, U.K. (e-mail: d.paccagnan@imperial.ac.uk).

Bolutife Ogunsula is with the Bloomberg LP, London EC4N4TQ, U.K. (e-mail: bogunsola@student.ethz.ch).

John Lygeros is with the Automatic Control Laboratory, ETH Zurich, 8092 Zurich, Switzerland (e-mail: lygeros@control.ee.ethz.ch).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2020.3044007.

Digital Object Identifier 10.1109/TAC.2020.3044007

3) Additionally, we show that classical algorithms for Nash equilibrium are not suitable for finding inertial Nash equilibria, as they violate the agents' rationality assumption by requiring them to perform detrimental moves. Motivated by these shortcomings, we propose a better-response dynamics algorithm, where agents switch action only if it is to their advantage when factoring the switching cost. We prove convergence to an inertial Nash equilibrium under weak assumptions (see Theorem 3).

We also position the concept in the context of existing literature, notably in relation to population games [11] and migration equilibria [12]. Furthermore, we show that introducing switching 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 convex.

*Organization:* In Section II, we introduce the notion of inertial Nash equilibrium, show its nonuniqueness and the nonconvexity 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 VI, study its monotonicity (more precisely, the lack thereof), and showcase how existing algorithms violate the agents' rationality assumption. In Section IV, we propose a better-response dynamics algorithm that provably converges to an inertial Nash equilibrium. Section V presents model extensions. In Section VI, we validate our model with a numerical study of taxi drivers' distribution in Hong Kong. Appendices A and B contain background material, and all the proofs.

*Notation:* The space of $n$-dimensional real vectors (resp. nonnegative, strictly positive) is denoted with $\mathbb{R}^n$ (resp. $\mathbb{R}^n_{\geq 0}$, $\mathbb{R}^n_{>0}$). The symbols $\mathbb{1}_n$ and $\mathbb{0}_n$ indicate the $n$-dimensional vector of unit entries and zero entries, respectively. 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 $\mathbf{e}_i$ denotes the $i$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 $x \neq \mathbb{0}_n$. $\|A\|$ is the induced 2-norm on $A$. Given $g(x) : \mathbb{R}^n \to \mathbb{R}^m$, 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. $\mathrm{Proj}_{\mathcal{X}}[x]$ 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 $u_i : \mathbb{R}^n_{\geq 0} \to \mathbb{R}$, 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 modeling 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.[1] This requirement can be formalized by introducing the unit simplex in dimension $n$, denoted with $\mathcal{S}$, and its relative interior $\mathcal{S}_+$

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

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

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

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 nonnegative coefficients $c_{ij}$ to represent the cost experienced by any agent when switching 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 switching to a different action, taking into account the cost of such switch.

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

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

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

*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. The following lemma shows that the set of inertial Nash equilibria contains the set of Nash equilibria, due to the nonnegativity of $c_{ij}$.

*Lemma 1:* A Nash equilibrium is an inertial Nash equilibrium.

The proof follows from Definitions 1 and 2, since condition (1) implies condition (2), as $c_{ij} \geq 0$ for all $i, j \in \{1, \dots, n\}$. In the following, we refer to an (inertial) Nash equilibrium as just an (inertial) equilibrium.

Despite being a natural extension to the traditional notions of equilibrium in game theory, to the best of authors' knowledge Definition 2 is novel. Its relevance stems from the observation that the coefficients $c_{ij}$ can model different and common phenomena, such as follows:

1) the tendency of agents to adhere to their habits, their reluctance to try something different, their loss aversion [13], or their risk aversion [14];
2) actual costs or fees that agents incur for switching action [15];
3) the lack of accurate information about other options [16].

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

[1] While we study the case where $\mathbb{1}_n^\top x = 1$, this is without loss of generality as identical results carry over to the case where $\mathbb{1}_n^\top x = \gamma > 0$.

*1) Area Coverage for Taxi Drivers:* Understanding and predicting the spatial distribution of taxi drivers has attracted the interest of the transportation community [17], [18]. 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 and on the fraction of drivers available in that location. The cost (fuel and time) that a driver incurs while moving between two different locations is captured by $c_{ij}$. Such model can help predict the driver distribution.

*2) Task Assignment in Server Network:* Consider a finite number of geographically dispersed servers connected through a network [19]. Each server corresponds to an action $i \in \{1, \ldots, n\}$. 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.

## B. Nonuniqueness and Nonconvexity 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

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

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

$$x_1 > 0 \quad \Rightarrow \quad \boxed{x_2 \geq x_1 - 0.2} \tag{3a}$$

$$x_1 > 0 \quad \Rightarrow \quad \boxed{x_2 \leq -2x_1 + 1.5} \tag{3b}$$

$$x_2 > 0 \quad \Rightarrow \quad x_2 \leq x_1 + 1 \tag{3c}$$

$$x_2 > 0 \quad \Rightarrow \quad x_2 \leq -0.5x_1 + 1 \tag{3d}$$

$$x_3 > 0 \quad \Rightarrow \quad \boxed{x_2 \geq -2x_1 + 1.1} \tag{3e}$$

$$x_3 > 0 \quad \Rightarrow \quad 2x_2 \geq -x_1 \tag{3f}$$

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

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 Fig. 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 Fig. 1. The points on this segment belong to the inertial equilibrium set even though they

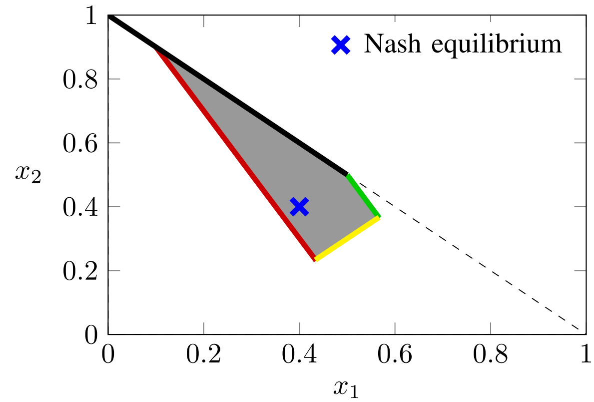


Fig. 1. Shaded region, including the thick red, yellow, green, and black lines, represents the inertial Nash equilibrium set for Example 1 projected on the plane $(x_1, x_2)$. The component $x_3$ can be reconstructed from $x_3 = 1 - x_1 - x_2$. The dashed line represents the simplex boundary, while the 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).

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 nonconvexity 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 nonconvex, even if we restrict attention to convex or concave utility functions.

*Remark 1:* While equilibrium uniqueness often makes the analysis simpler, this property does not always hold in real-world competitive decision making, see [20] for a discussion. Instead, convergence to one outcome or another often depends on the initial conditions and on the process utilized to revise the decisions [21], with examples ranging from road-traffic network to internet-routing. In this respect, it is worth remarking how celebrated notions, such as that of price of anarchy (the quality of the worst-performing equilibrium [22]), hinge precisely on the existence of multiple equilibria. Hence, multiplicity of equilibria should not be understood as an unwanted side-effect arising when accounting for switching costs, but instead as an element that enriches the model.

*Example 2:* In this example, we show that an inertial Nash equilibrium might exist even in absence of Nash equilibria. Consider a game with two actions and discontinuous utilities depicted in Fig. 2 as a function of $x_1$, since $x_1 + x_2 = 1$. The switching costs are $c_{11} = c_{22} = c_{21} = 0$, while $c_{12} > u_2(1 - \bar{x}_1) - u_1(\bar{x}_1)$. It is immediate to observe that no strategies $(x_1, x_2)$ constitute a Nash equilibrium, whereas $(\bar{x}_1, 1 - \bar{x}_1)$ is an inertial Nash equilibrium. We note that existence of Nash equilibria is guaranteed within our setting by merely assuming continuity of the utility functions (a direct application of the forthcoming Lemma 1 and Theorem 1).

## C. Related Work

The notion of inertial equilibrium is, to the best of authors' knowledge, novel, due to the presence of the switching costs $c_{ij}$. Similar models to that studied in this article arise in the

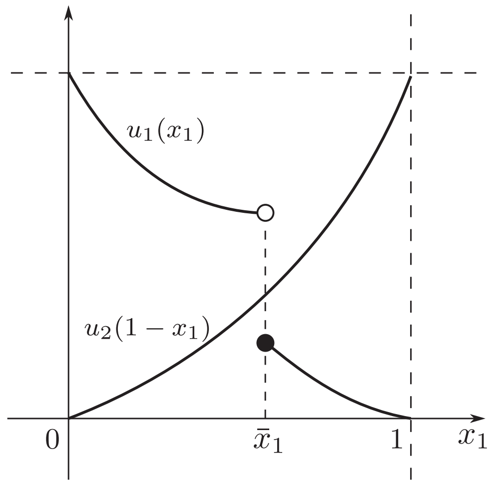


Fig. 2. Utilities of the game in Example 2.

context of *population games* (see [23], and references therein) with two important differences. First, we observe that in all these works switching cost are not accounted for, i.e., $c_{ij} = 0$, thus limiting the analysis only to equilibria in the sense of Definition 1 and not in the sense of Definition 2, which is instead the focus here. Second, the literature of population games focuses on the analysis and design of continuous-time agent dynamics that achieve an equilibrium in the sense of Definition 1, with different works providing local [11], [24] and global [25], [26] convergence guarantees. While a particular class of dynamics known as imitation dynamics are reminiscent of the discrete-time Algorithm 2 presented here, [11] and references therein provide convergence results to an equilibrium set (in the sense of Definition 1), whereas we provide convergence to a point in the inertial equilibrium set.

Additionally, we observe that mean-field games [27], [28] share with population games the fact that the agents are influenced by the overall population behaviour, and that both classes of games do not consider switching costs. We also note that within single-agent optimization formulations accounting for switching costs have been recently studied in [29] and [30].

A more closely related equilibrium concept was proposed in the study of migration models in the seminal works [12], [31] by Nagurney. These works introduce the notion of migration equilibrium in a way that resembles Definition 2, but with a number of important differences. Indeed, the problem formulation is different. In the migration equilibrium problem, we are given a fixed initial distribution $x^0 \in \mathcal{S}$, with $x^0_j$ 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 $(f_{ij})^n_{i,j=1}$ (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 $(f_{ij})^n_{i,j=1}$ 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. The difference between the two problems is fundamental. For instance, there exists a unique migration equilibrium[2] for the utilities of Example 1, which instead admits a set of several inertial equilibria. Finally, 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 (in the sense that the agents execute an algorithm to agree on the migrations and then carry out the agreed migrations).

## III. VI Reformulation

In this section, we first recall that the set of equilibria defined by (1) can be described as the solution of a certain VI. 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 definition of the VI 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 article are summarized in Appendix A.

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

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

The VI problem was first introduced in infinite dimensional spaces in [32], while the finite-dimensional VI in Definition 3 was studied for the first time in [33]. The monograph [34] includes a wide range of results on VI, amongst which their connection to Nash equilibria.

*Proposition 1 (Equilibria as VI solutions, [23, Th. 2.3.2]):* A point $\bar{x} \in \mathcal{S}$ is an equilibrium if and only if it is a solution of VI$(\mathcal{S}, -u)$, where $u(x) := [u_i(x)]^n_{i=1}$.

The following theorem shows that inertial equilibria can also be described by a VI.

*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 VI$(\mathcal{S}, F)$, where

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

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

Finally, we show that VI$(\mathcal{S}, F)$ reduces to VI$(\mathcal{S}, -u)$ in absence of switching costs, as one would expect.

*Lemma 2:* If $c_{ij} = 0$ for all $i, j$, then VI$(\mathcal{S}, F)$ is equivalent to VI$(\mathcal{S}, -u)$.

The migration equilibrium problem introduced in Section II-C also admits a VI characterization [35, eq. (5.11)], which is inherently different from VI$(\mathcal{S}, F)$. Indeed such VI admits one unique solution in the setup of Example 1 if $c_{ij}$ is a strictly

[2]Uniqueness holds for any initial distribution $x^0$, assuming that the migration cost $c_{ij}$ is a decreasing function of $f_{ij}$ for all $i, j$.

decreasing function of $f_{ij}$, while Fig. 1 shows that there are infinitely many inertial equilibria. The analogous of Lemma 2 holds also for the migration equilibrium VI, which reduces to VI($\mathcal{S}$, $-u$) if the migration costs $c_{ij}$ are zero.

### A. Lack of Monotonicity

If the operator $F$ in 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 [34, Ch. 12]). On the contrary, if this is not the case, the problem is known to be intractable in general, as nonmonotone variational inequalities supersede nonmonotone linear complementarity problems, which are known to be $\mathcal{NP}$-complete [36].

Since Proposition 6 in Appendix A, ensures that the solution set of a VI with monotone operator is convex, and since the inertial equilibrium set of Fig. 1 is not convex as explained in Section II, it follows that the corresponding VI operator $F$ cannot be monotone. The question is whether this observation extends to more general settings. In the following, we provide a strong negative result showing that the VI operator is *nonmonotone* in all nondegenerate instances of the inertial equilibrium problem where the utility functions are decreasing.

*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 there are no efficient algorithms to solve the VI problem [34]. The only technical assumption is that $\nabla_{x_i} u_i(x) < 0$, which is true for many applications; indeed the condition implies that $u_i(x)$ is decreasing in the number of agents on action $i$ increases, as commonly assumed in congestion problems [5], [37]. Moreover, the condition can be further weakened, as for the proof it suffices $\nabla_{x_{i^\star}} u_{i^\star}(x^\star) < 0$ only for specific $x^\star$, $i^\star$ 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 5 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. Existing Algorithms Violate the Rationality Assumption

Lemma 1 ensures that any equilibrium is an inertial equilibrium. Thus, one might try to use an algorithm for computing an equilibrium to determine an inertial equilibrium. 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 we generalize to other algorithms at the end of the section. We consider the projection algorithm [34, Alg. 12.1.1] for VI($\mathcal{S}$, $-u$), where $x(k)$ indicates the iterate $k$ of the algorithm.

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

*Proposition 2:* If $u_i$ is $L$-Lipschitz for all $i$, $\rho \leq 2/L$, and if there exists a concave function $\theta : \mathbb{R}^n \to \mathbb{R}$ 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.[3]

In the following, we analyze the behavior of Algorithm 1 on Example 1, and use it to highlight two 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 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 VI($\mathcal{S}$, $-u$), and thus, the unique equilibrium (see [34, Th. 2.3.3]). This is consistent with Lemma 1 and Fig. 1. 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 $x(0)$.

We now consider the following 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 switching 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, Algorithm 1 forces a mass of 0.1 to switch from action 3 to 2.

The drawbacks of Algorithm 1 are summarized as follows.

[3]The existence of a concave $\theta$ whose gradient matches $u(x)$ is guaranteed if $u_i$ depends only on $x_i$ 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 [34, Th. 12.1.11]. Finally, observe that if $-u$ is strongly monotone (see [34, Def. 2.3.1]), the projection algorithm converges without requiring the existence of $\theta(x)$ (see [34, Alg. 12.1.1]).

1) It violates the rationality assumption: agents are forced to switch action even when such switch is detrimental to their well being; this can even lead to forcing the agents to switch action when already at an inertial equilibrium.
2) The projection step requires the presence of a central operator. Such operator needs information not only on the utilities $u_i(x(k))$ for all $i$, but also on $x(k)$.

We note that drawback i) does not only apply to the specific Projection Algorithm 1, but rather to any algorithm, which solves VI $(\mathcal{S}, -u)$ (as, for example, the splitting methods and proximal-point methods reported in [34, Ch. 12]), because such VI does not account for the switching costs. The same is true for any population game algorithm [23], including the best-response dynamics and the replicator dynamics [25], as they attempt to solve a game with no notion of switching costs.

In the following section, we overcome these issues and present a better-response dynamics that

1) provably converges to an inertial equilibrium;
2) respects the agent's strategic nature;
3) requires limited coordination.

## IV. 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 switch to. The following fact immediately follows from Definitions 4 and 2 of inertial equilibrium.

*Lemma 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 $x_{i\to j}(k) \in [0, x_i(k)]$ switches 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 $x_{i\to j}(k) \in [0, x_i(k)]$
        $\Delta x_i(k) \leftarrow \Delta x_i(k) - x_{i\to j}(k)$,
        $\Delta x_j(k) \leftarrow \Delta x_j(k) + x_{i\to j}(k)$,
    **end repeat**
    $x(k+1) \leftarrow x(k) + \Delta x(k)$
    $k \leftarrow k + 1$

Different expressions of $x_{i\to j}(k)$ in Algorithm 2 give rise to a plethora of different agents' dynamics. This is analogous to evolutionary dynamics in population games, where the specific dynamics depend on the expression of the revision protocol [23, p. 121]. At this stage, we rather not give a particular expression to $x_{i\to j}(k)$, as the convergence of Algorithm 2 is guaranteed under very weak conditions and different choices of $x_{i\to j}(k)$. A possible modeling assumption sees agents switching from a less attractive action $i$ to a more favorable 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 $x_{i\to j}(k) = \beta x_i(k)$ with $\beta > 0$. A different modeling 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 and no agent switches if the current allocation is an inertial equilibrium. Second, 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 favorable action whenever one is available.

*Theorem 3 (Convergence of Algorithm 2):* Assume the following:

1) for each $i \in \{1, \dots, n\}$ the utility $u_i$ depends only on $x_i$, that $u_i$ is nonincreasing and $L$-Lipschitz;
2) there exists $c_{\min} > 0$ such that $c_{ij} \geq c_{\min}$ for all $i \neq j$ with $i, j \in \{1, \dots, n\}$;
3) there exist $0 < \tau \leq 1$, and $\varepsilon > 0$ such that at each iteration $k \in \mathbb{N}$, $x_{i\to j}(k) \geq 0$ for all $i \in \{1, \dots, n\}, j \in \mathcal{E}_i^{\text{out}}(x_k)$, and

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

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

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, see [38, eq. (2.1a)] for traffic networks, [1, eq. (4)] for plug-in electric vehicles, [39, eq. (3)] for taxi drivers. 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 $x_{i\to j}(k) \geq 0$ 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 nontrivial constraint imposed on $x_{i\to j}(k)$ is that on the left-hand side of (5a), and that of (5b); these are discussed in detail in Remark 2 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 2 (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 unfavorable action to a better one. Equation (5b), on the other hand, requires that only small fractions of the population switch action exactly at the same time. This is a realistic assumption for large populations, which are the focus of population games. Without condition (5b), 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 switch 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.

We point out that the computational complexity of Algorithm 2 depends on the particular choice of $x_{i\to j}$ (amount of agents moving to more promising actions). This is analogous to evolutionary dynamics in population games, where different choices of revision protocol give rise to a plethora of different dynamics, each presenting specific convergence rates [23, p. 121]. For this reason, an in-depth analysis of computational complexity of Algorithm 2 is beyond the scope of this work.

## V. Extensions

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

*Nonengaging 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 $\{1, \dots, n, e\}$. We 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. Introducing the additional action $e$ allows for agents to join and leave the game according to their interests.

For example, in the taxi area coverage 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 earlier 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 use Algorithm 2 by letting an agent $i$ switch to an arbitrary action whenever such action is attractive. A similar convergence result to that presented in Theorem 3 will hold. In particular, convergence is guaranteed if conditions (5a) and (5b) are satisfied, where we substitute the expression $\sum_{j\in\mathcal{E}_i^{\text{out}}(k)} x_{i\to j}(k)$ with the fraction of (discrete) agents that change action at time step $k$.

*Multiclass 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, in the spirit of [12]. 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 $u_i^\alpha : \mathbb{R}^n_{\geq 0} \to \mathbb{R}$, 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 *multiclass inertial equilibrium* if $\bar{x} \geq \mathbb{0}_{nA}$, $\mathbb{1}_n^\top \bar{x}^\alpha = \gamma^\alpha$ for all $\alpha$, and

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

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 $\mathcal{S} = \tilde{\mathcal{S}}^1 \times \cdots \times \tilde{\mathcal{S}}^A \subset \mathbb{R}^{nA}$ as the Cartesian product of the weighted simplexes $\tilde{\mathcal{S}}^\alpha = \{x^\alpha \in \mathbb{R}^n_{\geq 0},\ \mathbb{1}_n^\top x^\alpha = \gamma^\alpha\}$, one can redefine $F : \mathcal{S} \to \mathbb{R}^{nA}_{\geq 0}$, 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 multiclass inertial equilibria coincides with the solution set of VI $(\mathcal{S}, F)$. Theorem 2 about lack of monotonicity also extends to the multiclass 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 [17], [18], as it allows us to infer information for diverse scopes, including land-use classification and analysis of collective behavior of a city's population.

We focus on the urban area of Hong Kong, as the work [40] provides relevant data for our model. The authors of [40] divide the region of interest into $n = 18$ neighborhoods. We aim at determining an equilibrium distribution of the drivers across the different neighborhoods of the urban area, where each neighborhood corresponds to an action in the inertial 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

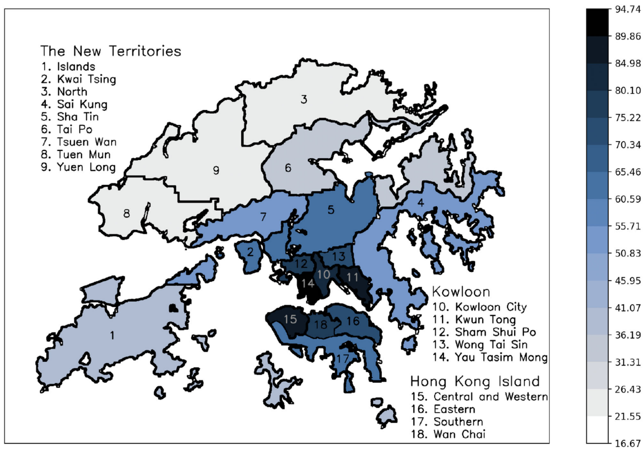


Fig. 3. Utility distribution at the equilibrium $\bar{x}$ achieved by Algorithm 2 with initial condition $x(0) = \mathbb{1}_n/n$. The values are expressed in HK\$ per hour. The map is that of [40, Fig. 1]. The central neighborhoods yield highest utilities, as one would expect.

neighborhood. This takes the form of

$$u_i(x_i) = \alpha_i t_i(x_i) - \kappa \tag{6}$$

where $\alpha_i$ is the average profit per trip in HK\$ as by [40, Fig. 4], and $t_i(x_i)$ is the expected number of trips per unit time, both relative to location $i$. The parameter $\kappa$ represents the operational cost per unit time.[4] The expression for $t_i(x_i)$ can be derived from the matching probability $m_i(x_i)$, which describes the probability that a taxi stationing in location $i$ is matched to a customer in a window of time of length $T$. According to both [39, eq. (4)] and [41, eq. (3)], we have

$$m_i(x_i) = 1 - e^{-\frac{p_i}{x_i x_{\mathrm{tot}}}}$$

where $x_{\mathrm{tot}}$ is the total number of drivers in the system, and customers are assumed to arrive in each window of length $T$ according to a Poisson distribution with rate $p_i$ [39, Sect. 3.1]. It follows[5] that the expected time before a taxi is matched equals $T/m_i(x_i)$, and therefore, a taxi is expected to complete a ride in an amount of time equal to $T/m_i(x_i) + D$, where $D$ is the average duration of a trip. Hence, the expected number of trips per unit time in location $i$ is $t_i(x_i) = (T/m_i(x_i) + D)^{-1}$, expression which fully characterizes the utility $u_i$ in (6). Through simple algebraic manipulations, $u_i$ can be shown to be nonnegative, nonincreasing for $x_i \geq 0$. We set $T = 1$ min, $x_{\mathrm{tot}} = 15\,333$ according to [40, Sec. 2.1], $D = 10$ min [40, Sec. 2.2], and choose $p_i$ to be proportional to the values in [40, Fig. 3], assuming a total number of daily passengers given by [42, Sec. 2]. With the chosen parameters, the Lipschitz constant of $\{u_i(x_i)\}_{i=1}^n$ with $x_i \in [0, 1]$ can be numerically found to be $L = 6.8 \cdot 10^3$. We let $c_{ij} = c_{ji}$ equal the fuel cost[6] of a trip from location $i$ to $j$, spread over an horizon of 15 min. In other words, a driver is willing to move from location $i$ to $j$, if after spending 15 min in the new location $j$, she would have made up a net profit at least as high as that in location $i$, plus the fuel cost.

[4]Since $\kappa$ is independent of the location, its value does not play a role in determining location equilibria. This is because $\kappa$ appears in both sides of (2) and, thus, cancels out.

[5]This follows from the assumption that being matched in a future time window is independent of being matched in the current time window.

[6]The fuel cost per km is given by [40, Sec. 2.2], the distances by [43].

TABLE I
ITERATIONS NEEDED FOR CONVERGENCE OF ALGORITHMS 1 AND 2 FOR DIFFERENT VALUES OF $\rho$ AND $\beta$

| Algorithm | # iterations mean | # iterations st. dev. |
|---|---|---|
| Alg. 1, $\rho = \rho^{\mathrm{theory}}$ | 7213 | 743 |
| Alg. 2, $\beta = \beta^{\mathrm{theory}}$ | 59 696 | 12 503 |
| Alg. 1, $\rho = 100 \cdot \rho^{\mathrm{theory}}$ | 150 | 6 |
| Alg. 2, $\beta = 100 \cdot \beta^{\mathrm{theory}}$ | 621 | 109 |
| Alg. 1, $\rho = 1000 \cdot \rho^{\mathrm{theory}}$ | does not converge | does not converge |
| Alg. 2, $\beta = 1000 \cdot \beta^{\mathrm{theory}}$ | 121 | 23 |

We report mean and standard deviation for 100 repetitions with random $x(0) \in \mathcal{S}$.

In our numerical study, we compare the projection Algorithm 1 with the better-response Algorithm 2, with stopping criterion $\|x(k+1) - x(k)\| \leq 10^{-6}$ and equal neighbor redistribution function $x_{i\to j}(k) = \beta x_i(k)$, $j \in \mathcal{E}_i^{\mathrm{out}}(k)$. For Algorithm 1, we choose a step-size $\rho^{\mathrm{theory}} = 1.5 \cdot 10^{-4}$ slightly smaller than $1/L$ to ensure convergence as by Proposition 2. The remaining assumptions of Proposition 2 hold as there exists a function $\theta(x)$ such that $\nabla_x \theta(x) = u(x)$ (see footnote 3). Similarly, to ensure convergence of Algorithm 2, we choose a stepsize $\beta^{\mathrm{theory}} = 2.4 \cdot 10^{-5}$ slightly smaller than $c_{\min}/L$ following the requirement[7] of Theorem 3.

Table I (top) shows that the number iterations required to reach convergence is substantial, due to the small values of $\rho$ and $\beta$ imposed by the theoretical convergence bounds. For this reason, we perform two additional simulations. We also perform additional simulations with larger step sizes for which convergence is not theoretically guaranteed. When step-sizes are multiplied by a factor of 100, both algorithms converge in all different repetitions with random initial conditions within much fewer iterations, as reported in Table I (mid). When the step-sizes are multiplied by a factor of 1000, Algorithm 1 fails to converge in all instances, while Algorithm 2 converges in all instances requiring only few iterations, as reported in Table I (bottom). This suggests that the theoretical bound obtained on $\beta$ is rather conservative when applied to the specific problem at hand. We also point out that each iteration of Algorithm 1 requires a projection, so it is more computationally expensive than one of Algorithm 2, which relies on additions and multiplications only.

## VII. CONCLUSION

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 VI, the resulting operator

[7]Ensuring $\beta < c_{\min}/L$ guarantees, in fact, that condition (5b) holds as $\sum_{i:j\in\mathcal{E}_i^{\mathrm{out}}(k)} x_{i\to j}(k) = \beta \sum_{i:j\in\mathcal{E}_i^{\mathrm{out}}(k)} x_i(k) \leq \beta < c_{\min}/L$.

is often nonmonotone. Thus, we proposed a natural dynamics that is distributed, and provably converges to an inertial Nash equilibrium. Future research directions include providing convergence rate guarantees for Algorithm 2 (for different choices of $x_{i\to j}$) and extending 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 VI that are used to characterize the equilibrium concepts introduced in Section II.

*Proposition 3 ([34, Prop. 2.3.3]):* Let $\mathcal{X} \subset \mathbb{R}^n$ be a compact, convex set and $F : \mathcal{X} \to \mathbb{R}^n$ be continuous. Then, VI$(\mathcal{X}, F)$ admits at least one solution.

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

*Proposition 4 ([34, Prop. 1.3.4]):* Assume that the set $\mathcal{X}$ can be described as $\mathcal{X} = \{x \in \mathbb{R}^n \,|\, g(x) \leq \mathbb{0}_m, h(x) = \mathbb{0}_p\}$, and that it satisfies Slater's constraint qualification in [44, eq. (5.27)]. Then, $\bar{x}$ solves 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 (7)

$$F(x) + \nabla_x g(x)\lambda + \nabla_x h(x)\mu = \mathbb{0}_n \tag{7a}$$

$$\mathbb{0}_m \leq \lambda \perp g(x) \leq \mathbb{0}_m \tag{7b}$$

$$h(x) = \mathbb{0}_p. \tag{7c}$$

We next recall the notion of monotonicity, which is a sufficient condition for convergence of a plethora of VI algorithms, see [34, Ch. 12].

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

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

*Proposition 5 (see [45, 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 semidefinite.

The definition of generalized Jacobian $\partial F(x)$ can be found in [46, 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 $(\nabla_x F(x) + \nabla_x F(x)^\top)/2 \succ 0$. We conclude this section with a result on the convexity of the VI solution set.

*Proposition 6 ([34, Th. 2.3.5]):* Let $\mathcal{X} \subseteq \mathbb{R}^n$ be closed, convex and $F : \mathcal{X} \to \mathbb{R}^n$ be continuous and monotone. Then, the solution set of VI$(\mathcal{X}, F)$ is convex.

## Appendix B
## Proofs

### A. Proof of Theorem 1

*Proof:* The proof consists in showing that the KKT system of VI$(\mathcal{S}, F)$ is equivalent to Definition 2 of inertial Nash. Since the set $\mathcal{S}$satisfies Slater's constraint qualification, by Proposition 4, VI$(\mathcal{S}, F)$ is equivalent to its KKT system

$$\begin{aligned} &F(x) + \mu\mathbb{1}_n - \lambda = \mathbb{0}_n \\ &\mathbb{0}_m \leq \lambda \perp x \geq \mathbb{0}_m \\ &\mathbb{1}_n^\top x = 1 \end{aligned} \tag{8}$$

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 above system can be compactly rewritten as

$$\mathbb{0}_n \leq \mu\mathbb{1}_n + F(x) \perp x \geq \mathbb{0}_n \tag{9a}$$

$$\mathbb{1}_n^\top x = 1. \tag{9b}$$

Observe that for any $x \in \mathcal{S}$ there exists $i^\star \in \{1, \ldots, n\}$ such that $F_{i^\star}(x) = 0$. Indeed, setting $i^\star \in \operatorname{argmax}_{i \in \{1,\ldots,n\}} u_i(x)$, gives $F_{i^\star}(x) = 0$ by the definition of $F$ in (4).

It follows that $\mu < 0$ is not possible, otherwise the nonnegativity 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 (9a), this would imply $x = \mathbb{0}_n$ thus violating (9b). We can conclude that $\mu = 0$ and (9) becomes

$$\begin{aligned} &\mathbb{0}_n \leq F(x) \perp x \geq \mathbb{0}_n \\ &\mathbb{1}_n^\top x = 1. \end{aligned} \tag{10}$$

System (10) is equivalent to $x \in \mathcal{S}$, and

$$x_i > 0 \underset{(10a)}{\Rightarrow} u_i(x) \geq u_j(x) - c_{ij} \qquad \forall i, j \in \{1, \ldots, n\}.$$

which coincides with Definition 2.

Existence of an inertial equilibrium follows readily from Proposition 3 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. ■

### B. Proof of Lemma 2

*Proof:* The vector $\bar{x}$ solves VI$(\mathcal{S}, F)$ if and only if

$$\sum_i \max_{j \in \{1,\ldots,n\}} (u_j(\bar{x}) - u_i(\bar{x}))(x_i - \bar{x}_i) \geq 0 \qquad \forall x \in \mathcal{S} \Leftrightarrow$$

$$\max_{j \in \{1,\ldots,n\}} u_j(\bar{x}) \underbrace{\sum_i (x_i - \bar{x}_i)}_{=0 \text{ as } x, \bar{x} \in \mathcal{S}} - \sum_i u_i(\bar{x})(x_i - \bar{x}_i) \geq 0 \qquad \forall x \in \mathcal{S}$$

which means, by definition, that $\bar{x}$ solves VI$(\mathcal{S}, -u)$. ■

### C. Proof of Theorem 2

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

1) We first show that there exists $\tilde{x} \in \mathcal{S}_+$ such that $\tilde{x}$ 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 $\lim_{m\to\infty} x(m) = \hat{x}$. 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

$$\begin{aligned}&\lim_{m\to\infty} u_i(x(m)) \geq \lim_{m\to\infty} u_j(x(m)) - c_{ij}\\ &\Leftrightarrow u_i(\hat{x}) \geq u_j(\hat{x}) - c_{ij}\end{aligned} \tag{11}$$

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

2) After establishing the existence of $\tilde{x} \in \mathcal{S}_+$, 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 $\mathcal{B}_{\tilde{\varepsilon}}(\tilde{x}) \cap \mathcal{S}_+$ is an inertial equilibrium. Let us reason again for the sake of contradiction. If for each $\varepsilon > 0$ there exists an inertial equilibrium in $\mathcal{B}_{\varepsilon}(\tilde{x}) \cap \mathcal{S}_+$, then we can construct a sequence of inertial equilibria converging to $\tilde{x}$. With the same continuity argument used in (11), 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 $\mathcal{B}_{\tilde{\varepsilon}}(\tilde{x}) \cap \mathcal{S}_+$ is an inertial equilibrium. By Rademacher's theorem [47, Th. 2.14], Lipschitzianity of $\{u_i\}_{i=1}^n$ guarantees[8] existence of $x^\star \in \mathcal{B}_{\tilde{\varepsilon}}(\tilde{x}) \cap \mathcal{S}_+$ such that $F$ is differentiable at $x^\star$.

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

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

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

$$u_{\ell_1}(x^\star) < u_{\ell_2}(x^\star) - c_{\ell_1\ell_2}. \tag{12}$$

Condition (12) is equivalent to $\ell_2 \in \mathcal{A}(\ell_1, x^\star)$ and $\ell_1 \notin \mathcal{A}(\ell_1, x^\star)$. If $\mathcal{A}(\ell_2, x^\star) = \{\ell_2\}$ then the statement is proven with $j^\star = \ell_1, i^\star = \ell_2$, otherwise there exists $\ell_3 \in \mathcal{A}(\ell_2, x^\star) \backslash \{\ell_2\}$. Note that it cannot be $\ell_3 = \ell_1$, because this means $u_{\ell_2}(x^\star) \leq u_{\ell_1}(x^\star) - c_{\ell_2\ell_1}$, which together with (12) results in $u_{\ell_1}(x^\star) < u_{\ell_1}(x^\star) - c_{\ell_2\ell_1} - c_{\ell_1\ell_2}$, 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 $\mathcal{A}(\ell_3, x^\star) = \{\ell_3\}$ then the statement is proven with $j^\star = \ell_2, i^\star = \ell_3$, otherwise there exists $\ell_4 \notin \{\ell_1, \ell_2, \ell_3\}$ such that $\ell_4 \in \mathcal{A}(\ell_3, x^\star)$. Since there are only $n$ different actions, by continuing the chain of reasoning, we conclude that there exists $k \in \{2, \ldots, n\}$ such that $\ell_k \in \mathcal{A}(\ell_{k-1}, x^\star)$ and $\mathcal{A}(\ell_k, x^\star) = \{\ell_k\}$, thus proving the statement with $j^\star = \ell_{k-1}$ and $i^\star = \ell_k$.

We now proceed to show that not only $i^\star \in \mathcal{A}(j^\star, x^\star)$, but actually $\mathcal{A}(j^\star, x^\star) = \{i^\star\}$. For the sake of contradiction, assume that there exists $\ell \neq i^\star$ such that $\ell \in \mathcal{A}(j^\star, x^\star)$. This means that $F_{j^\star}(x^\star) = u_{i^\star}(x^\star) - u_{j^\star}(x^\star) - c_{j^\star i^\star} = u_\ell(x^\star) - u_{j^\star}(x^\star) - c_{j^\star\ell}$. Then, consider the vector of the canonical basis $\mathbf{e}_{i^\star} \in \mathbb{R}^n$ and compute

$$\begin{aligned}&\lim_{t\to 0^+} \frac{F_{j^\star}(x^\star + t\mathbf{e}_{i^\star}) - F_{j^\star}(x^\star)}{t}\\ &= \lim_{t\to 0^+} \frac{[u_\ell(x^\star) - u_{j^\star}(x^\star) - c_{j^\star\ell}] - [u_\ell(x^\star) - u_{j^\star}(x^\star) - c_{j^\star\ell}]}{t}\\ &= 0\end{aligned} \tag{13}$$

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

$$\begin{aligned}&u_{i^\star}(x^\star + t\mathbf{e}_{i^\star}) - u_{j^\star}(x^\star) - c_{j^\star i^\star} < u_{i^\star}(x^\star) - u_{j^\star}(x^\star) - c_{j^\star i^\star}\\ &\quad = u_\ell(x^\star) - u_{j^\star}(x^\star) - c_{j^\star\ell}\end{aligned}$$

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

$$\begin{aligned}&\lim_{t\to 0^-} \frac{F_{j^\star}(x^\star + t\mathbf{e}_{i^\star}) - F_{j^\star}(x^\star)}{t}\\ &= \lim_{t\to 0^-} \frac{[u_{i^\star}(x^\star + t\mathbf{e}_{i^\star}) - u_{j^\star}(x^\star) - c_{j^\star i^\star}] - [u_{i^\star}(x^\star) - u_{j^\star}(x^\star) - c_{j^\star i^\star}]}{t}\\ &= \lim_{t\to 0^-} \frac{u_{i^\star}(x^\star + t\mathbf{e}_{i^\star}) - u_{i^\star}(x^\star)}{t} = \nabla_{x_{i^\star}} u_{i^\star}(x^\star) < 0\end{aligned} \tag{14}$$

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

$$\begin{aligned}&u_{i^\star}(x^\star + t\mathbf{e}_{i^\star}) - u_{j^\star}(x^\star) - c_{j^\star i^\star} > u_{i^\star}(x^\star) - u_{j^\star}(x^\star) - c_{j^\star i^\star}\\ &\quad = u_\ell(x^\star) - u_{j^\star}(x^\star) - c_{j^\star\ell}\end{aligned}$$

due to $\nabla_{x_{i^\star}} u_{i^\star}(x^\star) < 0$ by assumption. From (13) and (14), we obtain that $F_{j^\star}$ is not differentiable at $x^\star$, against what proved in the second part. Hence, we must conclude that there cannot exist $\ell \neq i^\star$ such that $\ell \in \mathcal{A}(j^\star, x^\star)$, thus, $\mathcal{A}(j^\star, x^\star) = \{i^\star\}$.

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

$$\begin{aligned}&u_{i^\star}(x^\star) - c_{j^\star i^\star} > u_\ell(x^\star) - c_{j^\star\ell} \qquad \forall\, \ell \neq i^\star\\ &u_{i^\star}(x^\star) - c_{i^\star i^\star} > u_\ell(x^\star) - c_{i^\star\ell} \qquad \forall\, \ell \neq i^\star.\end{aligned} \tag{15}$$

As a consequence of (15) there exists a small enough open ball around $x^\star$ where $F_{i^\star}(x^\star) = u_{i^\star}(x^\star) - u_{i^\star}(x^\star) - c_{i^\star i^\star} = 0$ and $F_{j^\star}(x^\star) = u_{i^\star}(x^\star) - u_{j^\star}(x^\star) - c_{j^\star i^\star}$. Thus

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

whose symmetric part has determinant $0 \cdot \nabla_{x_{j^\star}} u_{j^\star}(x^\star) - (\nabla_{x_{i^\star}} u_{i^\star}(x^\star))^2/4 < 0$, which makes $[\nabla_x F(x^\star)]_{i^\star j^\star \times i^\star j^\star}$ indefinite. Thus, $\nabla_x F(x^\star)$ itself is indefinite and $F$ is not monotone in $\mathcal{S}$due to Proposition 5. ■

### *D. Proof of Proposition 2*

*Proof:* Algorithm 1 is the projection algorithm in [34, Alg. 12.1.1], applied to VI($\mathcal{S}, -u$). A solution of VI($\mathcal{S}, -u$) exists by Proposition 3. The operator $-u$ is monotone in $\mathcal{S}$, because

[8]Rademacher's theorem assumes $F$ to be defined on an open subset of $\mathbb{R}^n$, but $\mathcal{S}_+$ is not open in $\mathbb{R}^n$. Indeed, one just needs to define $F$ on the $n-1$ dimensional open set $\{x \in \mathbb{R}^{n-1}_{>0} | \mathbb{1}^\top_{n-1} x < 1\}$, 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 $\{x \in \mathbb{R}^{n-1}_{>0} | \mathbb{1}^\top_{n-1} x < 1\}$, which implies existence of a differentiable point in the original $\mathcal{S}_+$.

$\theta$ is concave [48, eq. (12)]. Moreover, due to existence of $\theta$, $L$-Lipschitzianity is equivalent to $(1/L)$-cocoercitivity [49, Th. 18.15]. Then, for $\rho < 2/L$, Algorithm 1 is guaranteed to converge to a solution of VI($\mathcal{S}$, $-u$) by [34, Th. 12.1.8]. The final claim follows by observing that any Wardrop equilibrium is also an inertial Wardrop equilibrium (Lemma 1). ∎

***E. 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 $x_{i\to j}(k)$ 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

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

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

We now move our attention to proving the desired convergence statement. To do so, we will show that $x(k) \to \bar{x}$ such that $\mathcal{E}_i^{\text{out}}(\bar{x}) = \emptyset$ for all $i \in \{1,\dots,n\}$, thanks to the equivalence in Lemma 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 nondecreasing 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

$$\begin{aligned} u_i(k+1) - u_i(k) &\geq -L|x_i(k+1) - x_i(k)| \\ &\geq -L(c_{\min}/L - \varepsilon) = -c_{\min} + L\varepsilon =: -\gamma c_{\min} \end{aligned} \tag{16}$$

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

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 nonincreasing. Then there exists $j$ such that $i \in \mathcal{E}_j^{\text{out}}(x(k))$. It follows that

$$\begin{aligned} &i \text{ faces utility decrease at step } k \\ &\Rightarrow u_i(k) > u_j(k) + c_{ji} \geq \mu(k) + c_{\min}. \end{aligned} \tag{17}$$

Combining (16) with (17), we obtain

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

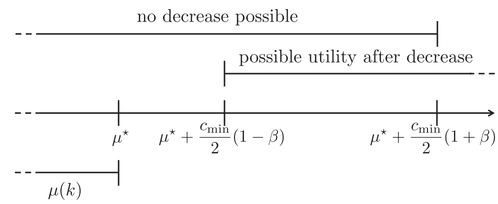


Fig. 4. Illustration of $\mu(k) \to \mu^\star$ from below and of inequalities (21) and (22) after iteration $\hat{k}$ (with $\gamma = 0.5$).

which implies $\mu(k+1) \geq \mu(k)$. Since $\mu(k)$ is nondecreasing and bounded ($\{u_i\}_{i=1}^n$ are continuous functions in a compact set), there exists a value $\mu^\star$ such that

$$\lim_{k\to\infty} \mu(k) = \mu^\star. \tag{18}$$

We show now that there exists an action $i^\star$ such that

$$\lim_{k\to\infty} u_{i^\star}(k) = \mu^\star. \tag{19}$$

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

$$\mu(k) > \mu^\star - c_{\min}(1-\gamma)/2 \quad \forall k \geq \hat{k}. \tag{20}$$

Then

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

where the first inequality follows from combining (17) and (20). Combining (16) and (21), we obtain

$$\begin{aligned} &i \text{ faces utility decrease at step } k \geq \hat{k} \\ &\Rightarrow u_i(k+1) \geq \mu^\star - c_{\min}(1-\gamma)/2 + c_{\min}(1-\gamma) \\ &= \mu^\star + c_{\min}(1-\gamma)/2. \end{aligned} \tag{22}$$

Fig. 4 illustrates inequalities (21) and (22).

Combining inequalities (21) and (22), we obtain that

$$\begin{aligned} &\exists\, k_1 \geq \hat{k} \text{ such that } u_i(k_1) \geq \mu^\star + \rho > \mu^\star \\ &\Rightarrow u_i(k) \geq \min\{\mu^\star + \rho, \mu^\star + c_{\min}(1-\gamma)/2\} \text{ for all } k \geq k_1. \end{aligned} \tag{23}$$

It then follows

$$\exists\, k_1 \geq \hat{k} \text{ such that } u_i(k_1) > \mu^\star \Rightarrow \lim_{k\to\infty} u_i(k) \neq \mu^\star. \tag{24}$$

By (24) and (18), it follows that there exists at least an action $i^\star$ such that $u_{i^\star}(k) \leq \mu^\star$ for all $k \geq \hat{k}$. Using again (18) and the "squeeze theorem" [50, Th. 3.3.6], we can conclude that $i^\star$ satisfies (19). 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 $\mathcal{E}_{i^\star}^{\text{in}}(x(k))$ is empty for $k \geq \hat{k}$ due to (17) and $u_{i^\star}(k) \leq \mu^\star$. In words, no other action can envy $i^\star$ after step $\hat{k}$. This implies that $u_{i^\star}(k)$ is a nondecreasing sequence, and in turn $x_{i^\star}(k)$ is a nonincreasing sequence. As a consequence

$$\lim_{k\to\infty} x_{i^\star}(k) = \bar{x}_{i^\star} \geq 0. \tag{25}$$

If $\bar{x}_{i^\star} = 0$, then clearly $\mathcal{E}^{\text{out}}_{i^\star}(\bar{x}_{i^\star}, x_{-i^\star}) = \emptyset$ by definition, for any $x_{-i^\star}$. If instead $\bar{x}_{i^\star} > 0$, since $x_{i^\star}(k+1) \leq (1-\tau)x_{i^\star}(k)$ due to (5a), then convergence is achieved in a finite number of steps. In other words, there exists $\tilde{k}$ such that $x_{i^\star}(k) = \bar{x}_{i^\star}$ for all $k \geq \tilde{k}$. In this case, for $k \geq \tilde{k}$ not only $\mathcal{E}^{\text{in}}_{i^\star}(x(k)) = \emptyset$, but also $\mathcal{E}^{\text{out}}_{i^\star}(x(k)) = \emptyset$, because otherwise $i^\star$ would encounter a mass decrease.

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

The last argument distinguishes two cases: $\bar{x}_{i^\star} > 0$ and $\bar{x}_{i^\star} = 0$. In the first case $\bar{x}_{i^\star} > 0$, we already showed that there exists $\tilde{k}$ such that $\mathcal{E}^{\text{in}}_{i^\star}(x(k)) = \mathcal{E}^{\text{out}}_{i^\star}(x(k)) = \emptyset$ for all $k > \tilde{k}$. Then, action $i^\star$ has no interaction with any the other action and considering $k \geq \tilde{k}$, we apply to $\{1,\ldots,n\}\backslash i^\star$, the previous reasoning until (25) to show that there is an action $j^\star \in \{1,\ldots,n\}\backslash\{i^\star\}$ with mass that converges to $\bar{x}_{j^\star}$ (in a finite number of steps if $\bar{x}_{j^\star} > 0$).

In the second case $\bar{x}_{i^\star} = 0$. Even though $\mathcal{E}^{\text{out}}_{i^\star}$ does not become the empty set at any finite iteration $k$, the mass $x_{i^\star}$ 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 $\eta(k) = \min_{j\in\{1,\ldots,n\}\backslash\{i^\star\}} u_j(k)$. Contrary to $\mu(k)$, the sequence $\eta(k)$ is not nondecreasing in general because the analogous of (17) does not hold, as action $i^\star$ could transfer some of its mass to $\{1,\ldots,n\}\backslash\{i^\star\}$ thus making their utilities decrease. Nonetheless, we show that there exists $\eta^\star$ such that

$$\lim_{k\to\infty} \eta(k) = \eta^\star. \tag{26}$$

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

$$x_{i^\star}(k) < \epsilon/(2L) \quad \forall\, k \geq k_\infty. \tag{27}$$

Let us now construct the sequence

$$\eta^0(k) = \eta(k) + \delta(k)$$

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

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

$$\lim_{k\to\infty} \delta(k) = \sum_{k=k_\infty}^{\infty} \max\{0, \eta(k) - \eta(k+1)\}$$

$$\leq L x_{i^\star}(k_\infty) \underset{(27)}{<} \epsilon/2$$

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

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

which proves (26).

Finally, we want to show that there exists $j^\star \in \{1,\ldots,n\}\backslash\{i^\star\}$ such that

$$\lim_{k\to\infty} u_{j^\star}(k) = \eta^\star. \tag{28}$$

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

$$\lim_{k\to\infty} u_\ell(x_\ell(k)) \neq \eta^\star. \tag{29}$$

Since $\eta(k) \to \eta^\star$, then $\max\{0, \eta(k) - \eta(k+1)\} \to 0$ as $k \to \infty$. This, together with $\eta(k) \to \eta^\star$, implies that condition (29) is equivalent to the existence of $\theta > 0$ such that for all $k' \geq 0$ there exists $k'' \geq k'$ such that

$$u_\ell(k'') > \eta^\star + \theta. \tag{30}$$

There are two possibilities in which $\ell$ can face a utility decrease after $k''$, namely through a mass transfer from some action $\{1,\ldots,n\}\backslash\{i^\star,\ell\}$ or through a mass transfer from action $i^\star$. If the mass transfer happens through some action $\{1,\ldots,n\}\backslash\{i^\star,\ell\}$, we can use the same argument of Fig. 4 and of implication (23) to conclude from (30) that

$$u_\ell(k) \geq \min\{\eta^\star + \theta, \eta^\star + c_{\min}(1-\gamma)/2\} \quad \forall\, k \geq k''. \tag{31}$$

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

$$x_{i^\star}(k) < \theta/(2L) \qquad \forall k \geq k' \tag{32}$$

and take $k''$ such that (30) holds. Then

$$u_\ell(k) \geq u_\ell(k'') - L\frac{\theta}{2L} > \eta^\star + \theta - \frac{\theta}{2} = \eta^\star + \frac{\theta}{2}. \tag{33}$$

for all $k \geq k''$, where the first inequality holds due to Lipschitz continuity and to (32), while the second inequality holds due to (30). We can conclude that if (29) holds for action $\ell$, then either (31) or (33) holds. Consequently, after $k''$ action $\ell$ does not attain the minimum $\eta(k)$. If (29) holds for all $\ell \in \{1,\ldots,n\}\backslash i^\star$, then the minimum $\eta(k)$ is not attained by any action after $k''$, which is a contradiction. Then, there must exist $j^\star$ such that (28) holds. With the same argument that led to (25), we can conclude that there exists $\bar{x}_{j^\star} \geq 0$ such that $\lim_{k\to\infty} x_{j^\star}(k) = \bar{x}_{j^\star} \geq 0$. As done for $i^\star$, we can conclude that $\mathcal{E}^{\text{out}}_{j^\star} = \emptyset$. ■

## References


[1] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 1, pp. 67–78, Jan. 2013.

[2] T. Alpcan, T. Başar, R. Srikant, and E. Altman, "CDMA uplink power control as a noncooperative game," *Wireless Netw.*, vol. 8, no. 6, pp. 659–670, 2002.

[3] R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Math. Oper. Res.*, vol. 29, no. 3, pp. 407–435, 2004.

[4] J. G. Wardrop, "Some theoretical aspects of road traffic research," *Proc. Institution 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, U.K.: Cambridge Univ. Press, 2007.

[6] F. Dieleman, *Households and Housing: Choice and Outcomes in the Housing Market*. Evanston, IL, USA: Routledge, 2017.

[7] S. Schulmeister, "A general financial transaction tax motives, revenues, feasibility and effects," *WIFO Studies*, *WIFO, number 31819,* Mar. Mar. 2008. [Online]. Available: https://www.wifo.ac.at/jart/prj3/wifo/main.jart?content-id=1454619331110&publikation_id=31819&detail-view=yes

[8] L. Coch and J. R. French Jr., "Overcoming resistance to change," *Hum. Relations*, vol. 1, no. 4, pp. 512–532, 1948.

[9] S. Oreg, "Resistance to change: Developing an individual differences measure," *J. Appl. Psychol.*, vol. 88, no. 4, pp. 680–693, 2003.

[10] O. Oyeniyi and A. Abiodun, "Switching cost and customers loyalty in the mobile phone market: The nigerian experience," *Bus. Intell. J.*, vol. 3, no. 1, pp. 111–121, 2010.

[11] W. H. Sandholm, "Potential games with continuous player sets," *J. Econ. Theory*, vol. 97, no. 1, pp. 81–108, 2001.

[12] A. Nagurney, "Migration equilibrium and variational inequalities," *Econ. Lett.*, vol. 31, no. 1, pp. 109–112, 1989.

[13] A. Tversky and D. Kahneman, "Loss aversion in riskless choice: A reference-dependent model," *Quart. J. Econ.*, vol. 106, no. 4, pp. 1039–1061, 1991.

[14] C. A. Holt and S. K. Laury, "Risk aversion and incentive effects," *Amer. Econ. Rev.*, vol. 92, no. 5, pp. 1644–1655, 2002.

[15] T. Zariphopoulou, "Investment-consumption models with transaction fees and Markov-chain parameters," *SIAM J. Control Optim.*, vol. 30, no. 3, pp. 613–636, 1992.

[16] P. Nelson, "Information and consumer behavior," *J. Political Econ.*, vol. 78, no. 2, pp. 311–329, 1970.

[17] P. S. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi GPS traces," in *Proc. Int. Conf. Perv. Comput.*, 2012, pp. 57–72.

[18] B. Li *et al.*, "Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset," in *Proc. IEEE Int. Conf. Perv. Comput. Commun. Workshops*, 2011, pp. 63–68.

[19] S. Abdallah and V. Lesser, "Learning the task allocation game," in *Proc. 5th Int. Joint Conf. Auton. Agents Multiagent Syst.*, 2006, pp. 850–857.

[20] J. Mercenier, "Nonuniqueness of solutions in applied general equilibrium models with scale economies and imperfect competition," *Econ. Theory*, vol. 6, no. 1, pp. 161–177, 1995.

[21] S. Dakhlia, "Testing for a unique equilibrium in applied general equilibrium models," *J. Econ. Dyn. Control*, vol. 23, no. 9/10, pp. pp. 1281–1297, 1999.

[22] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *Proc. Annu. Symp. Theor. Aspects Comput. Sci.*, 1999, pp. 404–413.

[23] W. H. Sandholm, *Population Games and Evolutionary Dynamics*. Cambridge, MA, USA: MIT Press, 2010.

[24] J. H. Nachbar, "Evolutionary selection dynamics in games: Convergence and limit properties," *Int. J. Game Theory*, vol. 19, no. 1, pp. 59–89, 1990.

[25] R. Cressman and Y. Tao, "The replicator equation and other game dynamics," *Proc. Nat. Acad. Sci.*, vol. 111, no. Suppl. 3, pp. 10810–10817, 2014.

[26] L. Zino, G. Como, and F. Fagnani, "On imitation dynamics in potential population games," in *Proc. IEEE 56th Annu. Conf. Decis. Control*, 2017, pp. 757–762.

[27] J.-M. Lasry and P.-L. Lions, "Mean field games," *Jpn. J. Math.*, vol. 2, pp. 229–260, 2007.

[28] M. Huang, P. E. Caines, and R. P. Malhamè, "Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized $\varepsilon$-Nash equilibria," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1560–1571, Sep. 2007.

[29] M. Lin, Z. Liu, A. Wierman, and L. L. H. Andrew, "Online algorithms for geographical load balancing," in *Proc. IEEE Int. Green Comput. Conf.*, *IEEE Comput. Soc.*, San Jose, CA, USA, Jun. 4–8, 2012, pp. 1–10.

[30] G. Goel and A. Wierman, "Online algorithm for smoothed regression and LQR control," in *Proc. 22nd Int. Conf. Artif. Intell. Statist.*, 2019, vol. 89, pp. 2504–2513.

[31] A. Nagurney, J. Pan, and L. Zhao, "Human migration networks," *Eur. J. Oper. Res.*, vol. 59, pp. 262–274, 1992.

[32] P. Hartman and G. Stampacchia, "On some non-linear elliptic differential-functional equations," *Acta Mathematica*, vol. 115, no. 1, pp. 271–310, 1966.

[33] R. W. Cottle, "Nonlinear programs with positively bounded Jacobians," *SIAM J. Appl. Math.*, vol. 14, pp. 147–158, 1966.

[34] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Berlin, Germany: Springer, 2007.

[35] A. Nagurney, *Network Economics: A Variational Inequality Approach*, vol. 10. Berlin, Germany: Springer, 2013.

[36] S.-J. Chung, "NP-completeness of the linear complementarity problem," *J. Optim. Theory Appl.*, vol. 60, no. 3, pp. 393–399, 1989.

[37] R. W. Rosenthal, "A class of games possessing pure-strategy Nash equilibria," *Int. J. Game Theory*, vol. 2, no. 1, pp. 65–67, 1973.

[38] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "Selfish routing in capacitated networks," *Math. Oper. Res.*, vol. 29, no. 4, pp. 961–976, 2004.

[39] N. Buchholz, "Spatial equilibrium, search frictions and dynamic efficiency in the taxi industry," Working Paper, Tech. Rep., 2019. [Online]. Available: https://scholar.princeton.edu/sites/default/files/nbuchholz/files/taxi_draft_2019_sm.pdf

[40] R. C. P. Wong, W. Y. Szeto, S. Wong, and H. Yang, "Modelling multi-period customer-searching behaviour of taxi drivers," *Transportmetrica B: Transport Dyn.*, vol. 2, no. 1, pp. 40–59, 2014.

[41] B. Petrongolo and C. A. Pissarides, "Looking into the black box: A survey of the matching function," *J. Econ. Literature*, vol. 39, no. 2, pp. 390–431, 2001.

[42] H. Yang, Y. W. Lau, S. C. Wong, and H. K. Lo, "A macroscopic taxi model for passenger demand, taxi utilization and level of services," *Transportation*, vol. 27, no. 3, pp. 317–340, 2000.

[43] Viamichelin, 2017. [Online]. Available: https://www.viamichelin.com

[44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[45] S. Dinh The Luc and S. Schaible, "Generalized monotone nonsmooth maps," *J. Convex Anal.*, vol. 3, pp. 195–206, 1996.

[46] F. H. Clarke, *Optimization and Nonsmooth Analysis*. Philadelphia, PA, USA: SIAM, 1990.

[47] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford, U.K.: Clarendon Press, 2000.

[48] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 35–49, May 2010.

[49] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Berlin, Germany: Springer, 2010.

[50] H. H. Sohrab, *Basic Real Analysis*. Berlin, Germany: Springer, 2003.



**Basilio Gentile** the master's degree in automation engineering from the University of Padova, Padua, Italy, and the master's degree in mathematical modeling and computation from the Technical University of Denmark, Technical University of Denmark, in 2013, and the Ph.D. degree in automatic control from the ETH Zürich, Zürich, Switzerland, in 2018.

In 2013, he spent seven months in the Motion Lab, University of California Santa Barbara, Santa Barbara, CA, USA, to work for his master's Thesis. His research focuses on aggregative games and network games.

**Dario Paccagnan** received the B.Sc. (Hons.) and M.Sc. (Hons.) degrees in aerospace engineering from the University of Padova, Padova, Italy, in 2011 and 2014, respectively, and the M.Sc. (Hons.) degree in mathematical modeling and computation from the Technical University of Denmark, Kongens Lyngby, Denmark, in 2014, and the Ph.D. degree in optimization and control from ETH Zurich, Zurich, Switzerland, in 2018.

He is an Assistant Professor (U.K. Lecturer) with the Department of Computing, Imperial College London, London, U.K., since the Fall 2020. Before that, he was a Postdoctoral Fellow with the Center for Control, Dynamical Systems and Computation, University of California, Santa Barbara, Santa Barbara, CA, USA. His research interests are at the interface between game theory and control theory, with a focus on the design of behavior-influencing mechanisms for socio-technical systems.

Dr. Paccagnan was recognized with the ETH Medal for his doctoral work, the SNSF Early Postdoc Mobility Fellowship, and the SNSF Doc Mobility Fellowship. He was also a finalist for the 2019 EECI Best Ph.D. Thesis Award.

**Bolutife Ogunsula** received the bachelor's degree in electrical and electronics engineering form the University of Lagos, Lagos, Nigeria, and the master's degree in robotics, systems and control engineering from ETH Zurich, Zurich, Switzerland, in 2012 and 2017.

He is a Software Engineer with the Bloomberg LP, New York, NY, USA. Prior to this he was was a Software Engineer for Codility.

**John Lygeros** (Fellow, IEEE) received the B.Eng. degree in electrical engineering and the M.Sc. degree in systems control, from the Imperial College of Science Technology and Medicine, London, U.K., in 1990 and 1991, respectively, and the Ph.D. degree in electrical engineering and computer sciences from the Electrical Engineering and Computer Sciences Department, University of California, Berkeley, Berkeley, CA, USA, in 1996.

During the period 1996–2000, he held a series of research appointments. Between 2000 and 2003, he was a University Lecturer with the Department of Engineering, University of Cambridge, Cambridge, U.K. Between 2003 and 2006, he was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Patras, Patras, Greece. In July 2006, he joined the Automatic Control Laboratory, 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 modeling, 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.

Dr. Lygeros is a member of the IET and the Technical Chamber of Greece.