Dissertation ETH Zurich No. 24975

Equilibria in aggregative games.

A thesis submitted to attain the degree of

Doctor of Sciences of ETH Zurich
(Dr. sc. ETH Zurich)

presented by

Basilio Gentile

Laurea Magistrale, Università di Padova, Italy
M.Sc., Technical University of Denmark
born on 22.08.1988 in Monselice
citizen of Italy

accepted on the recommendation of
Prof. Dr. John Lygeros (examiner)
Prof. Dr. Sandro Zampieri (co-examiner)
Prof. Dr. João Hespanha (co-examiner)

2018

Acknowledgments

I want to thank my doctoral advisor Prof. John Lygeros because of the trust, freedom and guidance that he gave me during the PhD. I always appreciated the acumen he brought to our discussions and admired his ability to understand a very wide range of technical research topics. I am grateful to Prof. Zampieri and Prof. Hespanha for carefully reading my dissertation and professionally refereeing my exam. Spending four years at IfA was an absolute privilege and I truly believe that its atmosphere of joyful hard work is unique, so I want to express my gratitude to all the people who created and shaped this magic place in the last decades. Carrying out everyday research with Francesca and Dario was energizing and stimulating; it is probably the first memory that comes to mind when thinking about the time here. I also want to thank Sergio for our joint efforts during the first part of my PhD, and the eight Master students that I had the honor to supervise. Teaching Advanced Topics in Control with Prof. Florian Dörfler turned out to be an absolute pleasure and I felt blessed for the responsibility of organizing and lecturing such a high-quality course. The list of the other IfA people who contributed to these amazing four years is too long, so I only mention my dear friends Marcello and Paul, buddies of several adventures across different continents. A special thought goes to my housemates Ricardo and Pedro, two exceptional mathematicians who have been important foundations of my Zurich life. I dedicate sincere words of gratitude to all teachers and mentors whose work sparked my curiosity and desire to learn, ranging from my elementary maths school teacher to my Master's Thesis advisor. I am deeply indebted to my mom and dad not only for their love and care, but also because the inevitable attraction towards completing a PhD is rooted in the importance they always devoted to knowledge and learning. Thanks from the bottom of my heart to my brother Filippo, who joined in Zurich soon after the beginning of my PhD and made my time much more fun and Zurich much more like home. I am lucky to have in my life Oktawia, who immediately understood the importance of this journey and walked along my side with support and dedication.

Abstract

This thesis studies equilibrium problems in aggregative games. A game describes the interaction among selfish rational agents, each of them choosing his strategy to optimize his own cost function, which depends also on the strategies of the other agents. In particular, the thesis focuses on aggregative games, where the cost of each agent is a sole function of his strategy and of the average agents' strategy. Not only such class of games can model a wide spectrum of applications, ranging from traffic or transmission networks to electricity or commodity markets, but it also lends itself to an elegant mathematical analysis.

The first part of the thesis investigates the relation between Nash and Wardrop equilibria, which are two classical concepts in game theory. Thanks to the powerful framework of variational inequalities, we derive bounds on the distance between the two equilibria and use them to show that the agents' strategies at the Nash equilibrium converge to those at the Wardrop equilibrium, when the number of agents grows to infinity. Moreover, we propose novel sufficient conditions to guarantee uniqueness of the Nash equilibrium for a specific aggregative game, which is often used in applications.

The second part of the thesis is dedicated to the design of algorithms that converge to Nash equilibrium and to Wardrop equilibrium in presence of constraints coupling the agents' strategies. Due to privacy issues and to the large number of agents at hand in real-life applications, centralized solutions might not be desirable. Hence, we first propose two parallel algorithms, where a central operator gathers and broadcasts aggregate information to coordinate the computations carried out by the agents. Then we design a distributed algorithm that only relies on local communications among the agents. We test the proposed algorithms in three case studies, where we also numerically verify the results of the first part of the thesis.

The last part of the thesis introduces the novel concept of equilibrium with inertia. Both classical Nash and Wardrop equilibria assume that each agent has the flexibility to change his strategy whenever this leads to an improvement. In some applications, however, this hypothesis is not realistic. We show that introducing an inertial coefficient which penalizes action switches leads to a richer set of equilibria, which is however in general not convex. Since classical algorithms for Nash and Wardrop equilibria cannot be used in presence of the inertial coefficients, we propose natural agents dynamics and guarantee their convergence to an equilibrium with inertial coefficients.

Sommario

Questa tesi studia problemi di equilibrio in giochi aggregativi. Un gioco descrive l'interazione fra agenti razionali, ognuno dei quali sceglie la sua strategia per ottimizzare la propria funzione di costo, che dipende anche dalle strategie degli altri agenti. In particolare, la tesi si concentra sui giochi aggregativi, in cui la funzione di costo di ogni agente dipende solo dalla sua strategia e dalla media delle strategie di tutti gli agenti. Questa classe di giochi si presta a modellare un ampio spettro di applicazioni, che includono reti di traffico e di trasmissione, o mercati di beni e dell'elettricità.

La prima parte della tesi analizza la relazione fra l'equilibrio di Nash e l'equilibrio di Wardrop, che sono due concetti classici in teoria dei giochi. Grazie al potente strumento delle disequazioni variazionali, deriviamo dei limiti sulla distanza fra i due equilibri, che vengono poi usati per mostrare che le strategie degli agenti all'equilibrio di Nash convergono alle strategie degli agenti all'equilibrio di Wardrop, quando il numero di agenti tende all'infinito. Inoltre, proponiamo nuove condizioni sufficienti per garantire l'unicità dell'equilibrio di Nash per uno specifico gioco aggregativo, che risulta molto rilevante nel contesto applicativo.

La seconda parte della tesi è dedicata allo sviluppo di algoritmi che convergono ad un equilibrio di Nash o ad un equilibrio di Wardrop in presenza di vincoli che accoppiano le stategie degli agenti. Per ragioni di privacy ed a causa dell'elevato numero di agenti in applicazioni tecnologiche, spesso le soluzioni centralizzate non sono praticabili. Per ovviare a ciò, proponiamo due algoritmi paralleli, in cui un operatore centrale raccoglie e trasmette informazioni aggregate per coordinare i calcoli portati avanti dagli agenti. Inoltre, presentiamo un algoritmo distribuito che utilizza soltanto comunicazioni locali fra gli agenti. Testiamo gli algoritmi presentati su tre applicazioni, grazie alle quali verifichiamo numericamente i risultati teorici della prima parte della tesi.

L'ultima parte della tesi introduce il nuovo concetto di equilibrio con inerzia. I concetti di equilibrio di Nash e di Wardrop assumono che ogni agente abbia la flessibilità per cambiare strategia ogni qual volta ciò porti ad un miglioramento. In alcune applicazioni, tuttavia, quest'ipotesi non è realistica. Mostriamo quindi che l'introduzione di un coefficiente d'inerzia, il quale penalizza cambi di azione, risulta in un insieme di equilibri che è più ricco ma in generale non convesso. Dato che i classici algortimi per l'equilibrio di Nash o di Wardrop non possono essere applicati in presenza dei coefficienti d'inerzia, proponiamo delle dinamiche che descrivono in modo naturale il comportamento degli agenti e mostriamo che convergono ad un equilibrio con coefficienti d'inerzia.

Contents

CHAPTER 1

Introduction

Classical problems in control and optimization are concerned with a single engineering system. The complexity of the decision making depends on the system itself and can rest on factors as diverse as nonlinearity, stochasticity, lack of stability or computational intractability. The perspective is however that of a single decision maker taking actions to achieve his goal; an example is an engineer who models and identifies a system to then design a stabilizing controller. Game theory, on the other hand, focuses on the interaction between different decision makers, or agents, each choosing his strategy to minimize his cost. Even though the individual agent can be described with a simple model, the interaction between different agents can give rise to complex behaviors. The difference between control/optimization and game theory is analogous to the difference between flying an airplane in the sky and optimally designing an urban road system. The complexity of the first task resides in understanding the vehicle dynamics and accounting for the weather conditions. The vehicle is just one, but a complicated one. The complexity of the second task lays in predicting how the human drivers respond to different road layouts. Each driver just takes the fastest route, but they all choose at the same time and influence each other. Since human decision makers play an important role not only in traffic networks, but indeed in many other modern technological and engineering systems, the understanding of game theory can be beneficial for analysis and design purposes.

Within the vast realm of game theory, this thesis focuses on the specific class of aggregative games, in which the cost of the individual agent only depends on his strategy and on some aggregate quantity, as for instance the average agents' strategy. Besides modeling a vast spectrum of applications, ranging from traffic or transmission networks to electricity or commodity markets, aggregative games lead to a simplification of the mathematical analysis, which can for instance be exploited to propose algorithms requiring little information exchange between the agents. Moreover, as aggregative games are characterized by an individual-aggregate interaction, they lend themselves particularly well to control purposes. In particular, it becomes natural to try and influence the aggregate strategy through macroscopic signals, which can be tolls in traffic networks or economic incentives in electricity markets.

To give an example of the modeling and control capabilities of aggregative games, let us consider the coordinated charging of electric vehicles (EVs). EVs are penetrating the market and their impact is foreseen to significantly grow in the next decades. They are in fact preferable over internal combustion engine vehicles because they present a lower cost per kilometer, do not pollute the air, and can provide ancillary services to the power grid. Based on these and other considerations, many countries are heavily subsidizing the EV sector. As of today the market penetration is relatively small, but different studies show that the effects of EVs charging on the power grid will reach a level that cannot be ignored. Hence it is necessary to coordinate the charging of the EVs themselves. However, this task is complicated because the decisions are ultimately taken by the EV owners, each deciding his charging schedule to save money on the electricity bill, without compromising habits and comfort. If the electricity price follows a dynamic scheme, i.e., it adapts based on the total EV consumption, then each agent's electricity bill only depends on his strategy and on the average agents' strategy. In order to design the dynamic price as a function of past and real-time measurements, it is crucial to understand that the aggregate EV response to such dynamic price is the result of interactions among selfish decision makers, which can be naturally modeled as a game. Moreover, the design task can rely on the aggregative nature of the interaction, that allows the price to be a sole function of the aggregate consumption, that is easily measurable with modern sensor technologies.

The first part of the dissertation focuses on the relation between Nash and Wardrop equilibria. The notion of *Nash equilibrium* is a central concept in game theory. It has been applied to a broad class of games in different fields such as economics, communication networks and electricity markets. Loosely speaking, a configuration constitutes a Nash equilibrium if no agent can improve his cost by unilaterally deviating from his strategy, considering the strategies of the other agents as fixed. However, in the context of aggregative games each agent's influence on the aggregate becomes negligible when the number of agents becomes very large. This consideration motivates the introduction of the *Wardrop equilibrium*, which describes a configuration where no agent can improve his cost by unilaterally deviating from his strategy, under the simplifying assumption that he has no influence at all on the aggregate. This concept is widely adopted in traffic network congestion under the name of user equilibrium, while it is often referred to as competitive equilibrium in the economics literature. The goal of the first part of the thesis is to provide a unifying framework for consistently studying the relations between Nash and Wardrop equilibria. Such framework helps in particular to derive bounds on the distance between the two equilibria as the number of agents grows large.

The second part of the thesis aims to design algorithms that achieve an equilibrium configuration, when the agents are not only coupled through their cost functions, but also through shared coupling constraints. Even though these constraints arise in many different applications, such as electricity markets or communication networks, many classical algorithms cannot be applied in their presence, as they crucially rely on the fact that the agents' strategy sets are decoupled. Regarding the information exchange, in the context of games centralized algorithms are often not feasible, because forcing the agents to compromise their privacy by communicating to a central coordinator their strategy sets and their cost functions is simply not an option. Moreover, the large number of agents at hand makes centralized computations of equilibria not tractable in many applications. For these reasons, we focus on the design of algorithms that present a parallel or distributed information structure. A *parallel algorithm* still features a central operator, but rather than collecting complete information from the agents and computing an equilibrium for them, he coordinates the execution of the algorithm by measuring and broadcasting aggregate quantities, such as the average agents' strategy. On the contrary, in a *distributed algorithm* each agent can only exchange information with his neighbors as specified by an underlying communication network.

The last part of the dissertation is perhaps the most innovative one, as it introduces the novel concept of *equilibrium with inertia*. Both classical Nash and Wardrop equilibria assume that each agent possesses the flexibility to change his strategy whenever this leads to an improvement. Such hypothesis is not necessarily realistic in some situations. Human decision makers are often prone to stick with their choices even though not optimal, because of lack of information about alternatives, physical or psychological burden of making a change, or even monetary transitional costs associated to the switch. Introducing costs for strategy changes leads to a richer set of equilibria, but classical algorithms for Nash and Wardrop equilibrium either lose guarantees of convergence, or present an execution that reveals to be detrimental for the agents. To overcome these issues, we tackle the algorithm design problem from the perspective of the agents and propose simple improvement dynamics in which agents switch strategy when this is actually beneficial for them, by taking into account also the cost of such strategy changes.

We do not provide a review of existing work in this introductory chapter because for each of our contributions throughout the thesis we conduct a specific literature comparison.

1.1 Outline and publications

Chapter 2 introduces the background material, comprising mathematical tools and results that are used throughout the thesis. In particular, we focus on the variational inequality (VI), which can be seen as a generalization of a convex optimization program and plays a fundamental role in the rest of the dissertation. All the material of Chapter 2 already exists in the literature.

Chapter 3 introduces aggregative games with coupling constraints along with the concepts of Nash equilibrium and Wardrop equilibrium. Based on a well-known reformulation of these equilibria as solutions of certain variational inequalities, we provide several sufficient conditions for uniqueness of Nash and of Wardrop; moreover, we bound the distance between Nash and Wardrop in terms of the number of agents in the game.

Chapter 4 focuses on the design of parallel algorithms for finding Nash and Wardrop equilibria. In particular, we propose a best-response algorithm, which is based on each agent computing his best possible strategy at every iteration, and a gradient-step algorithm, whose iteration requires each agent to perform a step in the steepest descent direction. In both cases, at every iteration the central coordinator measures the average agents' strategy along with the coupling constraint violation and broadcasts primal and dual aggregate quantities to the agents in order to guarantee convergence.

Chapters 3 and 4 are based on the articles

[GPP17] B. Gentile*, F. Parise*, D. Paccagnan*, M. Kamgarpour, and J. Lygeros, *Nash and Wardrop equilibria in aggregative games with coupling constraints*, arXiv preprint arXiv:1702.08789 (2017),

[PGP16] D. Paccagnan*, B. Gentile*, F. Parise*, M. Kamgarpour, and J. Lygeros, *Distributed computation of generalized Nash equilibria in quadratic aggregative games with affine coupling constraints*, Proceedings of the IEEE Conference on Decision and Control (2016), IEEE, pp. 6123–6128,

where a star indicates equal author contributions. In particular the article [GPP17] is an extension of the preliminary work [PGP16].

Chapter 5 proposes a gradient-step distributed algorithm for computation of Nash and Wardrop equilibria in presence of coupling constraints. We also prove a standalone result on parametric variational inequalities, which indeed is crucial to show convergence of the distributed algorithm. Chapter 5 is based on the following article:

[PGP16] F. Parise*, B. Gentile*, and J. Lygeros, *A distributed algorithm for average aggregative games with coupling constraints*, arXiv preprint arXiv:1706.04634 (2017).

Chapter 6 studies three specific applications. The first two are the charging of electric vehicles and the route choice game in a congested traffic network; they demonstrate and numerically verify the results presented in Chapters 3 and 4. The third case study is a Cournot game with transportation costs, which validates via simulation the theoretical findings of Chapter 5.

Chapter 7 introduces the novel concept of Wardrop equilibrium with inertia. After investigating its relations with the classical Wardrop equilibrium, we equivalently characterize the set of Wardrop equilibria with inertia as the solution set of a variational inequality. Finally, we propose agents dynamics that converge to a Wardrop equilibrium with inertia. The results in this chapter are based on the work

[PGP16] B. Gentile, D. Paccagnan, B. Ogunsola, and J. Lygeros, *A novel concept of equilibrium over a network*, Proceedings of the IEEE Conference on Decision and Control (2017), IEEE, pp. 6123–6128.

Even though the material of Chapters 3-7 is extracted from the aforementioned four articles, the presentation in this dissertation introduces more examples and provides interpretations, reports proofs that are omitted in the articles and conducts more detailed comparisons with the literature. In particular Chapter 7 contains substantial amount of results that are not present in [GPO17]. In the rest of the dissertation we report our minor findings as lemmas, our major contributions as theorems and the results of other authors as propositions. In the following we present the notation used throughout the thesis.

1.2 Notation

Vectors

The space of n-dimensional real vectors is denoted with $\mathbb{R}^n$, while $\mathbb{R}^n_{\geq 0}$ is the space of non-negative n-dimensional real vectors and $\mathbb{R}^n_{>0}$ 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 $\mathbf{e}_i$ denotes the i^{th} vector of the canonical basis. Given $x^1, \dots, x^M$, with $x^i \in \mathbb{R}^n$ for all i, we use $\{x^i\}_{i=1}^M$ to denote the set of M vectors, $(x^i)_{i=1}^M$ to denote the sequence of vectors, and $[x^i]_{i=1}^M = [x^1; \dots; x^M] = [(x^1)^\top, \dots, (x^M)^\top]^\top \in \mathbb{R}^{Mn}$ to denote the stacked vector. Moreover, we indicate $x^{-i} = [x_1; \dots; x_{i-1}; x_{i+1}; \dots; x_M] \in \mathbb{R}^{(M-1)n}$. The symbol $\|x\|$ denotes the 2-norm of $x \in \mathbb{R}^n$.

Matrices

Given $A \in \mathbb{R}^{n \times n}$, $A \succ 0$ $(\succeq 0) \Leftrightarrow x^\top A x = \frac{1}{2} x^\top (A + A^\top) x > 0$ $(\geq 0), \forall x \neq \mathbb{0}_n$. Thus when checking positive (semi) definiteness of a matrix one has to consider its symmetric part. $\mathrm{diag}(A)$ is the diagonal matrix which has the same diagonal of A. $\mathrm{blkdiag}(A_1, \dots, A_M)$ is the block diagonal matrix whose blocks are the matrices $A_1, \dots, A_M$. $\|A\|$ 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 $[\nabla_x g(x)]_{i,j} := \frac{\partial g_j(x)}{\partial x^i}$. Given $g(x) : \mathbb{R} \to \mathbb{R}$, we denote $g'(x) = \frac{\partial g(x)}{\partial x}$. I_n denotes the $n \times n$ identity matrix and $A \otimes B$ denotes the Kronecker product.

Sets

Given the sets $\mathcal{X}^1, \ldots, \mathcal{X}^M \subseteq \mathbb{R}^n$, we denote

$$\frac{1}{M}\sum_{i=1}^{M} \mathcal{X}^i = \{z \in \mathbb{R}^n \,|\, \exists (x^1, \ldots, x^M) \text{ such that } x^i \in \mathcal{X}^i, \forall\, i \text{ and } z = \frac{1}{M}\sum_{i=1}^{M} x^i\}.$$

Given $\mathcal{X} \subseteq \mathbb{R}^n$ and the function $f : \mathcal{X} \to \mathbb{R}^n$, we say that f is continuous if it is continuous in its domain $\mathcal{X}$. We use $\mathcal{X}^{-i} = \mathcal{X}^1 \times \ldots \mathcal{X}^{i-1} \times \mathcal{X}^{i+1} \times \ldots \mathcal{X}^M$. The convex hull of $\mathcal{X}^1, \ldots, \mathcal{X}^M$ is denoted with $\operatorname{conv}(\mathcal{X}^1, \ldots, \mathcal{X}^M)$. All the sets are assumed to be non-empty.

Miscellaneous

$\operatorname{Proj}_{\mathcal{X}}[x]$ is the Euclidean projection of the vector x onto the set $\mathcal{X}$. The symbol $:=$ means "equal by definition". The notation $\bar{x} \in \operatorname{argmin}_{\mathcal{X}} f(x)$ indicates that $\bar{x}$ belongs to the set of minimizers of f over $\mathcal{X}$. When the minimizer is unique, we sometimes use the notation $\bar{x} = \operatorname{argmin}_{\mathcal{X}} f(x)$. $\mathcal{U}[a, b]$ represents the uniform distribution on the real interval $[a, b]$. All the definitions that are not referenced can be found in at least one of the books [FP03, Ber07, BC10].

CHAPTER 2

Mathematical background

The goal of this chapter is to introduce the mathematical concepts that are needed for the results presented in Chapters 3-6. All the definitions, statements and proofs already exist in the literature. We here present and summarize concepts appearing in different works according to our perspective, with the attempt of highlighting connections and differences.

2.1 Basics of variational inequalities

In this section we present the variational inequality, which is an important mathematical tool used used throughout the thesis.

Definition 1 (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 variational inequality VI$(\mathcal{X}, F)$ if

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

The variational inequality (VI) problem was first introduced in infinite dimensional spaces by the mathematician Guido Stampacchia in the 1960s [HS66], in the context of boundary problems defined by partial differential equations. The finite-dimension VI problem of Definition 1 was identified and studied for the first time in the PhD thesis of Richard Cottle [Cot66], under the supervision of George Dantzig, well-known for fundamental contributions in linear programming and operations research in general.

Indeed the VI problem is strictly related to mathematical programming. To shed light on the connection between the two problem classes, let us consider the optimization program

$$\underset{x \in \mathcal{X}}{\operatorname{argmin}}\, f(x), \tag{2.1}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ and $f : \mathcal{X} \to \mathbb{R}$. The following proposition introduces a well-known first-order optimality condition for (2.1).

Proposition 1 ([BT97, p. 210])**.** Assume that the function f is continuously differentiable on the closed[1], convex set $\mathcal{X} \subseteq \mathbb{R}^n$.

1. Any local minimizer $\bar{x}$ of f must satisfy the first-order optimality condition
$$\nabla_x f(\bar{x})^\top (x - \bar{x}) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.2}$$
2. If f is convex on $\mathcal{X}$, then any $\bar{x}$ satisfying (2.2) is a global minimizer for f. □

As for some other statements of this chapter, the proof is not reported and it can be found in the reference provided. The solutions of (2.2) are called stationary points of f. Since (2.2) is equivalent to VI($\mathcal{X}$,∇f) by Definition 1, the problem of finding the stationary points can be cast as a variational inequality problem. The next proposition clarifies that the opposite does not hold in general, i.e., not every variational inequality problem can be cast as a stationary point problem.

Proposition 2. [OR00, Proposition 4.1.6] Let the operator $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on the closed, convex set $\mathcal{X}$. There exists a function $f : \mathcal{X} \to \mathbb{R}$ such that $\nabla_x f(x) = F(x)$ for all $x \in \mathcal{X}$ (in words, F is a gradient-operator) if and only if the Jacobian matrix $\nabla_x F(x)$ is symmetric for all $x \in \mathcal{X}$. □

In particular, if we assume the function f in (2.1) to be convex on $\mathcal{X}$, then by Proposition 1 the program (2.1) is equivalent to the stationary point problem (2.2). Hence a convex optimization program can be seen as a specific instance of a variational inequality, as visually represented in Figure 2.1.

The KKT system

In the following we review the role of the Karush-Kuhn-Tucker (KKT) system in optimization programs and variational inequalities. To this end, let us focus on a constraint set $\mathcal{X}$ of the form
$$\mathcal{X} = \{x \in \mathbb{R}^n \,|\, g(x) \leq \mathbb{0}_m, h(x) = \mathbb{0}_p\}, \tag{2.3}$$
with $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$, or in components $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, p$. The KKT system relative to (2.1) reads [Ber99, p. 310]
$$\nabla_x f(x) + \nabla_x g(x)\lambda + \nabla_x h(x)\mu = \mathbb{0}_n \tag{2.4a}$$
$$\mathbb{0}_m \leq \lambda \perp g(x) \leq \mathbb{0}_m \tag{2.4b}$$
$$h(x) = \mathbb{0}_p, \tag{2.4c}$$

[1]Throughout the rest of the thesis, we say that a function is differentiable on a closed set $\mathcal{X} \neq \mathbb{R}^n$ if there exists an open superset of $\mathcal{X}$ where the function is differentiable.

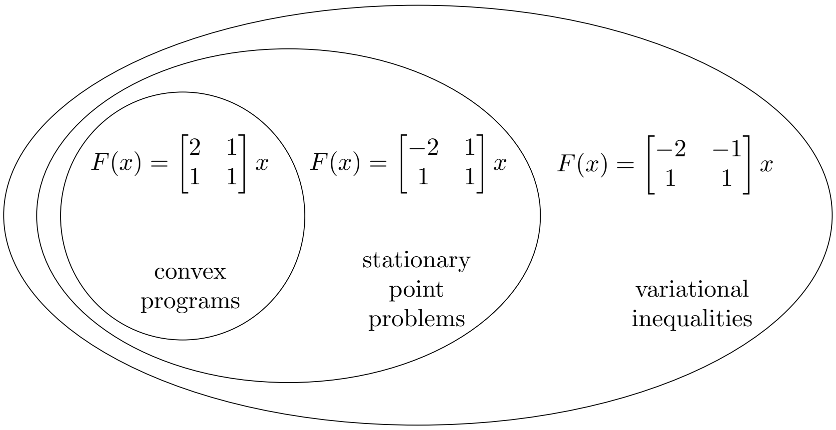


Figure 2.1: Inclusions among different problem classes. The three sample VI problems share the same convex, closed set $\mathcal{X} \subseteq \mathbb{R}^2$ and feature linear operators. The sample optimization program is $\underset{x\in\mathcal{X}}{\operatorname{argmin}}\, 2x_2^1 + 2x_1x_2 + x_2^2$, while the sample stationary point problem consists in finding the stationary points of $-2x_1^2 + 2x_1x_2 + x_2^2$ over $\mathcal{X}$. The left matrix is symmetric and positive definite, the middle matrix is symmetric but indefinite, the right matrix is asymmetric and indefinite.

where $\lambda \in \mathbb{R}^m$ is the dual variable associated with the constraint $g(x) \leq \mathbb{0}_m$ and $\mu \in \mathbb{R}^p$ is the dual variable associated with the constraint $h(x) = \mathbb{0}_p$.

We introduce Slater's constraint qualification, which is then used to establish the connection between the program (2.1) and the KKT system (2.4).

Definition 2 (Slater's constraint qualification [BV04, eq. (5.27)])**.** The set $\mathcal{X}$ in (2.3) is said to satisfy Slater's constraint qualification if the function $g_i : \mathbb{R}^n \to \mathbb{R}$ is convex for all $i = 1, \dots, m$, the function $h_j : \mathbb{R}^n \to \mathbb{R}$ is affine for all $j = 1, \dots, p$, the gradients $\{\nabla_x h_j\}_{j=1}^p$ are linearly independent, and there exists $x \in \mathcal{X}$ such that $g_i(x) < 0$ for all $i = 1, \dots, m$. □

Proposition 3 ([PAE13, Theorems 5.33 and 5.45])**.** Assume that the set $\mathcal{X}$ satisfies Slater's constraint qualification and that f is differentiable in $\mathcal{X}$.

1. If $\bar{x}$ solves the program (2.1), then there exist $\bar{\lambda}$, $\bar{\mu}$ such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ solves the KKT system (2.4).

2. Assume that f is convex in $\mathcal{X}$. If $(\bar{x}, \bar{\lambda}, \bar{\mu})$ solves the KKT system (2.4), then $\bar{x}$ solves the program (2.1). □

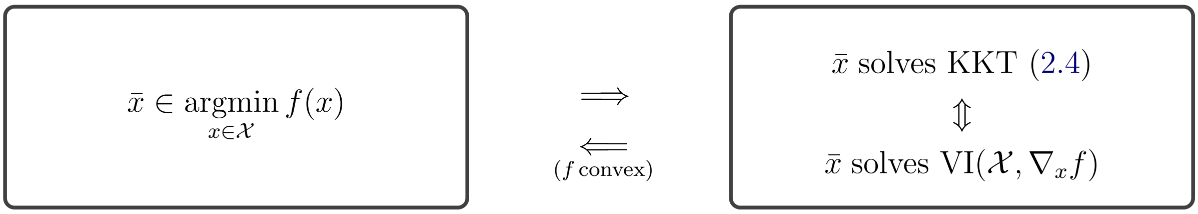


Figure 2.2: Relation between (convex) optimization program, KKT system and variational inequality, under Slater's constraint qualification. With $\bar{x}$ solution of KKT (2.4) we mean that there exist $\bar{\lambda}$, $\bar{\mu}$ such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ solves (2.4).

The KKT system relative to VI($\mathcal{X}$,F) reads

$$F(x) + \nabla_x g(x)\lambda + \nabla_x h(x)\mu = \mathbb{0}_n \tag{2.5a}$$
$$\mathbb{0}_m \leq \lambda \perp g(x) \leq \mathbb{0}_m \tag{2.5b}$$
$$h(x) = \mathbb{0}_p, \tag{2.5c}$$

where the only difference with (2.4) is that $\nabla_x f(x)$ is replaced by $F(x)$, just as $\nabla_x f(x)$ in (2.2) is replaced by $F(x)$ in Definition 1. We are now ready to draw the connection between VI($\mathcal{X}$,F) and the KKT system (2.5). We include the proof, as it is short and instructional.

Proposition 4 ([FP03, Proposition 1.3.4])**.** Assume that the set $\mathcal{X}$ satisfies Slater's constraint qualification. 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 (2.5).

Proof. By Definition 1, $\bar{x}$ solves VI($\mathcal{X}$,F) if and only if

$$\bar{x}^\top F(\bar{x}) \leq x^\top F(\bar{x}), \ \forall\, x \in \mathcal{X} \quad \Leftrightarrow \quad \bar{x} \in \operatorname*{argmin}_{x \in \mathcal{X}} \ x^\top F(\bar{x}). \tag{2.6}$$

Since $\mathcal{X}$ satisfies Slater's constraint qualification and $x^\top F(\bar{x})$ is convex, differentiable in x, by Proposition 3 the program (2.6) is equivalent to its KKT system, that reads as (2.5). □

Propositions 3 and 4 can be made more general by replacing Slater's constraint qualification with the less stringent Abadie constraint qualification, as in [PAE13, Theorems 5.33] and [FP03, Proposition 1.3.4]. We choose to use Slater's constraint qualification because it is satisfied in the applications of Sections 6.1, 6.2, 6.3 and because it helps keeping the exposition simpler[2], thus allowing to focus on the relations between VI($\mathcal{X}$,F) and the KKT system, which is summarized in Figure 2.2.

In a nutshell, the important concepts presented so far are that

[2] Verifying Abadie constraint qualification requires knowledge of the solution $\bar{x}$, hence Slater's constraint qualification is easier to guarantee a priori.

1. VI($\mathcal{X}$,$\nabla_x f$) and the KKT system (2.4) are first-order necessary conditions for the program (2.1), which become sufficient if f is convex,

2. the KKT system (2.5) of a VI is a straightforward generalization of the KKT system (2.4) of an optimization program, and

3. variational inequalities generalize convex optimization programs.

Our interest in variational inequality problems lies in the fact that not only they generalize optimization programs, but they are also equivalent characterizations of many equilibrium problems of interest, as it will be explained in Chapters 3 and 7.

Existence and uniqueness of VI solutions

The program (2.6) constitutes an equivalent reformulation of VI($\mathcal{X}$,F). We now introduce a second equivalent reformulation of VI($\mathcal{X}$,F) as fixed point problem, which is used in this subsection to derive results on existence and uniqueness of VI solutions, and in Section 2.3 to lay the foundations for algorithms to find solution of the VI.

Proposition 5. Consider a closed, convex set $\mathcal{X} \subseteq \mathbb{R}^n$ and a scalar $\tau > 0$. Then

$$\bar{x} \text{ solves } \mathrm{VI}(\mathcal{X}, F) \Leftrightarrow \bar{x} = \operatorname*{Proj}_{\mathcal{X}}[\bar{x} - \tau F(\bar{x})]. \tag{2.7}$$

Proof. By definition of projection,

$$\begin{aligned} \bar{x} = \operatorname*{Proj}_{\mathcal{X}}[\bar{x} - \tau F(\bar{x})] &\Leftrightarrow \bar{x} = \operatorname*{argmin}_{x \in \mathcal{X}} (x - \bar{x} + \tau F(\bar{x}))^\top (x - \bar{x} + \tau F(\bar{x})) \\ &\overset{\text{Prop. 1}}{\Leftrightarrow} 2(x - \bar{x} + \tau F(\bar{x}))^\top_{|x=\bar{x}} (y - \bar{x}) \geq 0, \quad \forall y \in \mathcal{X} \\ &\Leftrightarrow F(\bar{x})^\top (y - \bar{x}) \geq 0, \quad \forall y \in \mathcal{X}, \end{aligned}$$

which yields (2.7) by Definition 1. □

Proposition 6 ([FP03, Proposition 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. The solution set of VI($\mathcal{X}$,F) is non-empty.

Proof. The operator $\bar{x} \to \operatorname*{Proj}_{\mathcal{X}}[\bar{x} - \tau F(\bar{x})]$, which is from the compact, convex set $\mathcal{X}$ into itself, is continuous, because F and the projection operator are continuous. By Brouwer's fixed point theorem [FP03, Theorem 2.1.18], such operator admits a fixed point. The conclusion then follows from the equivalence (2.7). □

To state the main result about uniqueness of the solution of a VI, we introduce the concept of strong monotonicity.

Definition 3 (Monotone and strongly monotone)**.** An operator $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is α-strongly monotone (i.e., strongly monotone with monotonicity constant $\alpha > 0$) if

$$(F(x) - F(y))^\top (x - y) \geq \alpha \|x - y\|^2, \tag{2.8}$$

for all $x, y \in \mathcal{X}$. The operator is monotone if (2.8) holds for $\alpha = 0$. □

Proposition 7 ([FP03, Theorem 2.3.3])**.** Let $\mathcal{X}$ be closed, convex and $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be continuous and strongly monotone. Then VI($\mathcal{X}$,F) admits one unique solution. □

To verify whether an operator is strongly monotone or monotone one can exploit the following equivalent characterization.

Proposition 8. [FP03, Proposition 2.3.2] Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex. A continuously differentiable operator $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is strongly monotone with monotonicity constant $\alpha > 0$ if and only if

$$\nabla_x F(x) \succeq \alpha I_n, \qquad \forall x \in \mathcal{X}. \tag{2.9}$$

It is monotone if and only if

$$\nabla_x F(x) \succeq 0, \qquad \forall x \in \mathcal{X}. \tag{2.10}$$

Moreover, if $\mathcal{X}$ is compact then there exists $\alpha > 0$ such that $\nabla_x F(x) \succeq \alpha I_n$ for all $x \in \mathcal{X}$ if and only if $\nabla_x F(x) \succ 0$ for all $x \in \mathcal{X}$. □

As specified in the notation section, conditions (2.9) and (2.10) are to be read $(\nabla_x F(x) + \nabla_x F(x)^\top)/2 \succeq \alpha I_n$ and $(\nabla_x F(x) + \nabla_x F(x)^\top)/2 \succeq 0$, i.e., positive (semi) definiteness is expressed on the symmetric part.

Proposition 8 sheds light on the relations between convexity and monotonicity, which are summarized in the following Table 2.1.

$f(x)$		$\nabla_x f(x)$
convex	$\Leftrightarrow$	monotone
strongly convex	$\Leftrightarrow$	strongly monotone

Table 2.1: Equivalence between monotonicity and convexity for a twice continuously differentiable function f.

To prove the equivalences of Table 2.1, note that by Proposition 8 the operator $\nabla_x f(x)$ is strongly monotone if and only if the Hessian matrix satisfies $\nabla_x \nabla_x f(x) \succeq \alpha I_n$ for all $x \in \mathcal{X}$, which is equivalent [BV04, eq. (9.7)] to strong convexity of f. In the same way one can show that $\nabla_x f(x)$ is monotone if and only if f is convex [SPF10, eq. (12)].

In view of Table 2.1, we can conclude that Proposition 7 generalizes to VIs the fact that a strongly convex function over a convex, closed set admits a unique minimizer [FP03, p. 78].

We conclude with a trivial lemma, which we nonetheless report because it is often referred to in the remaining chapters.

Lemma 1. If $F : \mathcal{X} \to \mathbb{R}$ is α-strongly monotone and $G : \mathcal{X} \to \mathbb{R}$ is monotone, then $F + G$ is α-strongly monotone. If F and G are monotone, then $F + G$ is monotone.

Proof. The first statement is proven by

$$\begin{aligned} &(F(x) + G(x) - F(y) - G(y))^\top (x - y) \\ &= \underbrace{(F(x) - F(y))^\top (x - y)}_{\geq \alpha \|x-y\|^2} + \underbrace{(G(x) - G(y))^\top (x - y)}_{\geq 0} \geq \alpha \|x - y\|^2. \end{aligned} \tag{2.11}$$

The second statement is proven by replacing α with 0 in (2.11). □

2.2 Operator properties

In the previous section we introduced monotonicity and α-strong monotonicity. We now present and analyze other two properties of $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ that are used in the thesis. The properties of Definitions 3-5 will then serve in the next Section 2.3 as sufficient conditions for convergence of two simple algorithms to a solution of VI($\mathcal{X}$,F).

Definition 4 (Lipschitz, nonexpansive, contractive)**.** An operator $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz continuous (with constant $L > 0$) if

$$\|F(x) - F(y)\| \leq L\|x - y\|$$

for all $x, y \in \mathcal{X}$. F is nonexpansive if it is Lipschitz with constant $L = 1$. F is δ-contractive if it is Lipschitz with constant $\delta \in [0, 1)$. □

Definition 5 (Cocoercive)**.** The operator $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is η-cocoercive (i.e., cocoercive with cocoercitivity constant $\eta > 0$) if for all $x, y \in \mathcal{X}$

$$(F(x) - F(y))^\top (x - y) \geq \eta ||F(x) - F(y)||^2. \qquad \square$$

We point out that Definitions 3, 4 and 5 are provided with the standard dot product and corresponding 2-norm, but they could be expressed with any scalar product in $\mathbb{R}^n$ and corresponding norm. While Definitions 4-5 are already sufficient to move on with the following material, the rest of Section 2.2 is dedicated to introducing a visual tool to relate the different operator properties. Given Definitions 3, 4 and 5, one could ask questions like:

1. can a cocoercive operator be contractive?

2. does cocoercitivity imply strong monotonicity?

The following Figure 2.3 helps answering these questions.

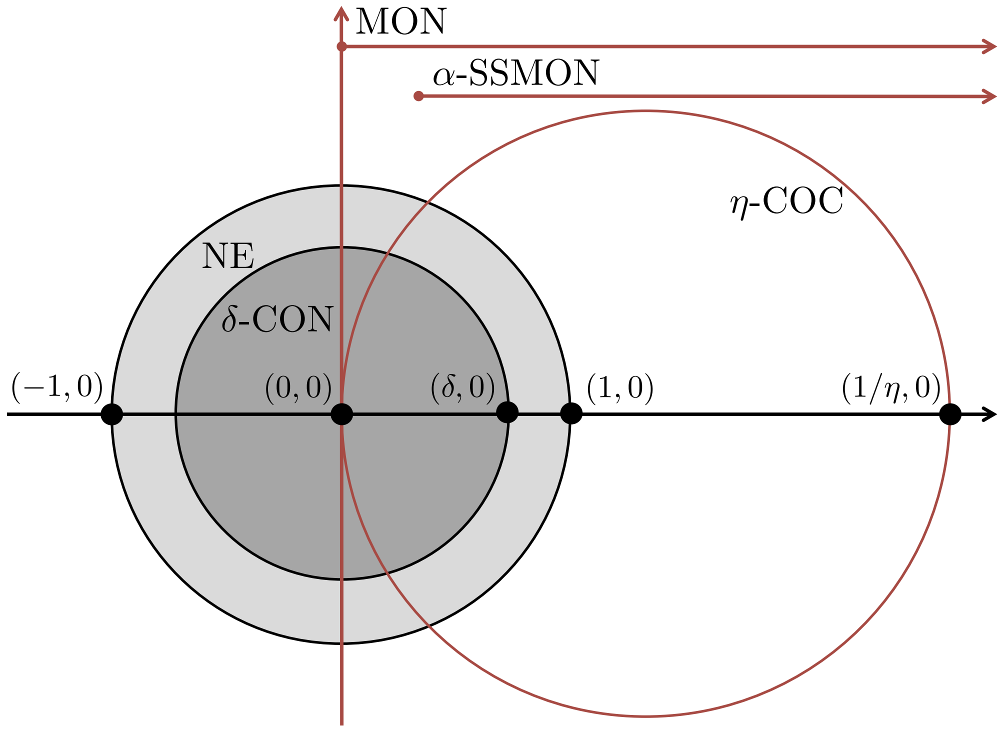


Figure 2.3: Illustration in $\mathbb{R}^2$ of δ-contractiveness (δ-CON), nonexpansiveness (NE), α-strong monotonicity (α-SSMON), η-cocoercitivity (η-COC), monotonicity (MON). For each of these five properties we draw the corresponding region where $F(1,0)$ is restricted to be, assuming that $(0,0)$ is a fixed point of F.

We derive the analytical expression of the regions in Figure 2.3. Let us denote the two Cartesian coordinates of $F(1,0)$ as $(p,q) = F(1,0)$.

- Strongly monotone and monotone:

$$((p,q)-(0,0))^\top((1,0)-(0,0)) \geq \alpha\|(1,0)-(0,0)\| \Leftrightarrow p \geq \alpha.$$

- Lipschitz, contractive and nonexpansive:

$$\|(p,q)-(0,0)\| \leq L\|(1,0)-(0,0)\| \Leftrightarrow p^2+q^2 \leq L^2.$$

- Cocoercive:

$$((p,q)-(0,0))^\top((1,0)-(0,0)) \geq \eta\|(p,q)-(0,0)\|^2$$
$$\Leftrightarrow p \geq \eta(p^2+q^2) \Leftrightarrow p^2 - \frac{p}{\eta} + \frac{1}{4\eta^2} + q^2 \leq \frac{1}{4\eta^2} \Leftrightarrow \left(p-\frac{1}{2\eta}\right)^2 + q^2 \leq \left(\frac{1}{2\eta}\right)^2.$$

The idea of Figure 2.3 comes from [GB14, Figure 1]. While it clearly fails to completely characterize the operators (for instance, it assumes that one fixed point exists, which is not always the case), it provides the visual intuition to relate the different operator properties. For instance, the two questions above can be answered by inspecting Figure 2.3.

1. Nothing prevents a cocoercive operator to be contractive and the COC region is contained in the CON region when $\eta > 1$, implying that an operator η-cocoercive with $\eta > 1$ is also contractive. Indeed this last claim can be proved analytically:

$$\|F(x)-F(y)\|^2 \overset{\text{COC}}{\leq} \frac{1}{\eta}(F(x)-F(y))^\top(x-y) \overset{\substack{\text{Cauchy}\\ \text{Schwarz}}}{\leq} \frac{1}{\eta}\|F(x)-F(y)\|\|x-y\|$$
$$\Rightarrow \|F(x)-F(y)\| \leq \frac{1}{\eta}\|x-y\|.$$

2. By inspecting Figure 2.4, one can guess that there can be a cocoercive operator which is not strongly monotone. Indeed an example is $F(x) = \mathbb{0}_n$ for all x. In the same way, Figure 2.4 hints that any strongly monotone and Lipschitz operator is cocoercive. Indeed, this is proved by

$$(F(x)-F(y))^\top(x-y) \overset{\substack{\text{strong}\\ \text{mon.}}}{\geq} \alpha\|x-y\|^2 \overset{\text{Lipschitz}}{\geq} \frac{\alpha}{L}\|F(x)-F(y)\|^2. \quad (2.12)$$

With the goal of providing a visual reference more complete than Figure 2.3 for relating different operator properties, we introduce in the following some further definitions, which are typically used to guarantee convergence of certain algorithms to a fixed point, as in [Ber07, Theorem 2.1], [CP02, p. 522], or in [GPC16, Corollary 1] for game theoretical applications. Such definitions are not used in the rest of the thesis and jumping to Section 2.3 does not affect the understanding of the following chapters.

Definition 6 (Pseudocontractive, firmly nonexpansive)**.** An operator $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is said to be

a. γ-strongly-pseudocontractive (with constant $\gamma > 0$) if $I_n - F$ is γ-strongly monotone.

b. pseudocontractive if $I_n - F$ is monotone.

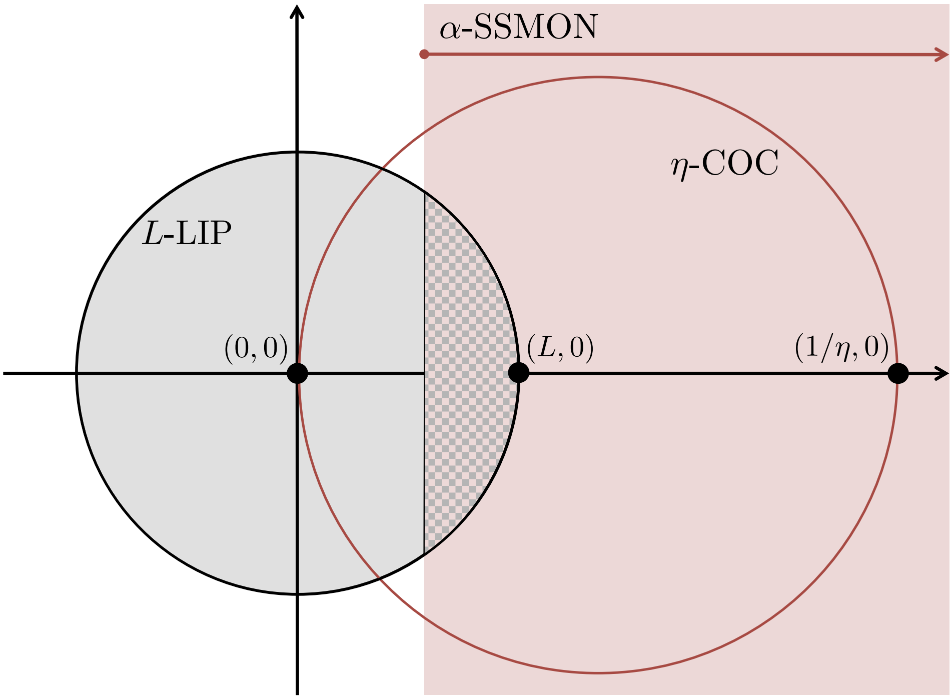


Figure 2.4: In gray the Lipschitz region, in brown the strongly-monotone region, in gray-brown squares their intersection, which is contained in a cocoercive ball for η small enough.

c. ρ-strictly-pseudocontractive (with constant $\rho < 1$) if

$$\|F(x) - F(y)\|^2 \leq \|x - y\|^2 + \rho\|F(x) - F(y) - (x - y)\|^2 \tag{2.13}$$

d. firmly non-expansive if (2.13) holds with $\rho = -1$.

Proposition 9. The operator $F : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is η-cocoercive if and only if the operator $I_n - F$ is $(1 - 2\eta)$-strictly pseudo contractive.

Proof. We prove the statement by equivalently rewriting $(1 - 2\eta)$-strict pseudocontractiveness of $I_n - F$ as η-cocoercitivity of F:

$$\|F(x) - F(y) - (x - y)\|^2 \leq \|x - y\|^2 + (1 - 2\eta)\|F(x) - F(y)\|^2$$
$$\Leftrightarrow \|F(x) - F(y)\|^2 - 2(F(x) - F(y))^\top (x - y) + \|x - y\|^2 \leq \|x - y\|^2 + \|F(x) - F(y)\|^2$$
$$- 2\eta\|F(x) - F(y)\|^2$$
$$\Leftrightarrow -2(F(x) - F(y))^\top (x - y) \leq -2\eta\|F(x) - F(y)\|^2$$
$$\Leftrightarrow (F(x) - F(y))^\top (x - y) \geq \eta\|F(x) - F(y)\|^2.$$
□

In Figure 2.5 we complete the visual representation of Figure 2.3 by including in cyan the regions relative to the properties of Definition 6, whose analytical derivation is straightforward given Definition 6 and Proposition 9. Indeed, the way we realized validity of

Proposition 9 was by first drawing Figure 2.5 based on Definition 6 and then equating the radii of the two circles, i.e., imposing $1 + \frac{1+\rho}{1-\rho} = \frac{1}{\eta}$, which yields $\rho = 1 - 2\eta$.

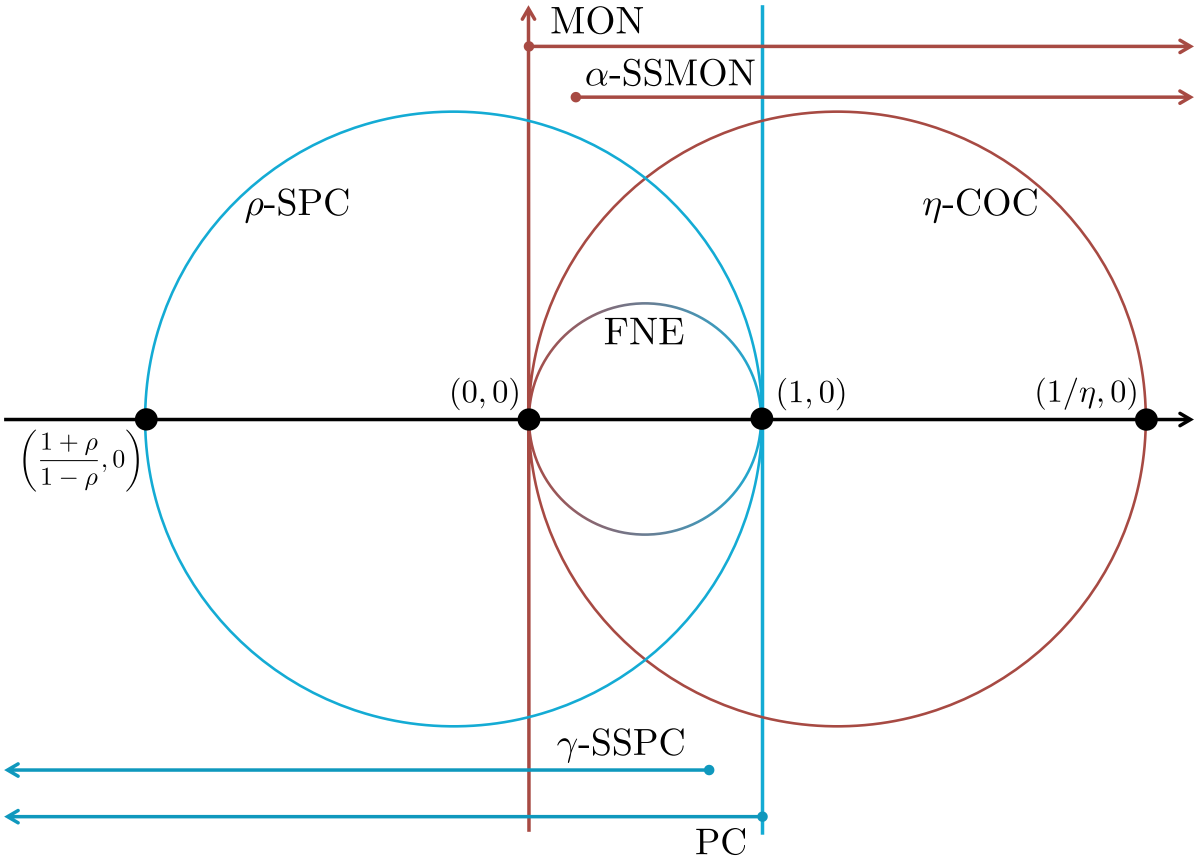


Figure 2.5: Illustration in $\mathbb{R}^2$ of some of the properties of Figure 2.3 along with γ-strong-pseudocontractiveness (γ-SSPC), ρ-strict-pseudocontractiveness (ρ-SPC), pseudocontractiveness (PC).

Figure 2.5 suggests that firm-nonexpansiveness, which is equivalent to (-1)-strict pseudocontractiveness, can also be regarded as 1-cocoercitivity; this can be easily verified analytically.

Moreover, as a consequence of Proposition 9, there is a one-to-one relation between the properties of Definitions 3-5 and those of Definition 6, as detailed in the following Table 2.2. Such relation results in the symmetric role of the properties in brown and cyan of Figure 2.5.

F		$I_n - F$
strongly monotone	$\Leftrightarrow$	strongly pseudocontractive
cocoercive	$\Leftrightarrow$	strictly pseudocontractive
monotone	$\Leftrightarrow$	pseudocontractive

Table 2.2: Relations between operator properties.

2.3 Algorithms for variational inequalities

In this section we study two algorithms for finding a solution of VI($\mathcal{X}$,F). We start with the simple projection algorithm with constant step length τ, which generalizes to VI the gradient projection algorithm [Ber99, eq. (2.31)] for optimization programs.

The projection algorithm

Algorithm 1: Projection algorithm

Initialization $\tau > 0$, $k = 0$, $x_{(0)} \in \mathcal{X}$

Iterate $x_{(k+1)} = \operatorname{Proj}_{\mathcal{X}} \left[x_{(k)} - \tau F(x_{(k)}) \right]$

$k \leftarrow k + 1$

In the rest of this section we consider $\mathcal{X}$ to be closed, convex, F to be Lipschitz and study convergence of Algorithm 1 under further assumptions on F. By Definitions 3 and 5 and by (2.12) it follows that

$$F \text{ strongly monotone} \Rightarrow F \text{ cocoercive} \Rightarrow F \text{ monotone},$$

as can be also seen by looking at Figures 2.3 and 2.4.

Proposition 10 ([FP03, Theorem 12.1.8]). Let $\mathcal{X} \subseteq \mathbb{R}^n$ be closed, convex, let $F : \mathcal{X} \to \mathbb{R}^n$ be η-cocoercive and let VI($\mathcal{X}$,F) admit a solution. Then for any $x_{(0)}$, the sequence $(x_{(k)})_{k=0}^{\infty}$ generated by Algorithm 1 with $\tau < 2\eta$ converges to a solution of VI($\mathcal{X}$,F). □

By Proposition 10, under Lipschitz continuity both cocoercitivity and strong monotonicity are sufficient to guarantee convergence of Algorithm 1. The next example shows how Lipschitz continuity and monotonicity are not sufficient to guarantee convergence of Algorithm 1.

Example 1 ([FP03, Example 12.1.3]). Consider the closed and convex set $\mathbb{R}^n$. By Definition 1, for any operator $F : \mathbb{R}^n \to \mathbb{R}^n$, the solution of VI($\mathbb{R}^n$,$F$) coincides with the solution of the system of equations $F(x) = \mathbb{0}_n$. Let us define the linear operator $F : \mathbb{R}^2 \to \mathbb{R}^2$ as

$$F(x_1, x_2) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{2.15}$$

The unique solution of VI($\mathcal{X}$,F) is $\mathbb{0}_n$. Monotonicity of the operator F is guaranteed by

$$\nabla_x F(x) = \frac{1}{2}\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \succeq 0, \qquad \forall\, x \in \mathcal{X},$$

together with Proposition 8. Linearity implies Lipschitz continuity. Let us now study the sequence produced by Algorithm 1. For any $\tau > 0$ we have

$$x_{(k+1)} = \operatorname*{Proj}_{\mathbb{R}^2} [x_{(k)} - \tau F(x_{(k+1)})] = x_{(k)} - \tau F(x_{(k+1)}) = \begin{bmatrix} 1 & -\tau \\ \tau & 1 \end{bmatrix} x_{(k)}. \tag{2.16}$$

The eigenvalues of the matrix in (2.16) are $1 \pm i\tau$, where i is the imaginary unit, and they are both outside of the unit circle. We can conclude that Algorithm 1 does not converge to the solution of VI($\mathcal{X}$,F), unless initialized at the solution. □

The operator in (2.15) is Lipschitz and monotone, but it is not a gradient operator, i.e., there exists no function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $\nabla f(x) = F(x)$ for all x, as can be seen by Proposition 2 and the fact that the matrix in (2.15) is not symmetric. This is crucial to show that Algorithm 1 does not converge. Indeed, if the operator F is Lipschitz, monotone and it is a gradient operator, then convergence of Algorithm 1 is guaranteed by the following Proposition 11 and Corollary 1.

Proposition 11 ([BC10, Theorem 18.15])**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be closed, convex and $f : \mathcal{X} \to \mathbb{R}$ be convex, differentiable in $\mathcal{X}$. Then the gradient $\nabla f : \mathcal{X} \to \mathbb{R}^n$ is L-Lipschitz if and only if it is $1/L$-cocoercive. □

Corollary 1. Let $\mathcal{X} \subseteq \mathbb{R}^n$ be closed, convex, $F : \mathcal{X} \to \mathbb{R}^n$ be monotone, L-Lipschitz, and VI($\mathcal{X}$,F) admit a solution. Assume that F is a gradient operator, i.e., there exists a function $f : \mathcal{X} \to \mathbb{R}$ such that $\nabla f(x) = F(x)$ for all x in $\mathcal{X}$. Then for any $x_{(0)}$, the sequence $(x_{(k)})_{k=0}^{\infty}$ generated by Algorithm 1 with $\tau < 2/L$ converges to a solution of VI($\mathcal{X}$,F).

Proof. Since ∇f is monotone, then f is convex (see Table 2.1) and we can use Proposition 11 to conclude that $F = \nabla f$ is $1/L$ cocoercive. Convergence of Algorithm 1 for $\tau < 2\eta = 2/L$ is then guaranteed by Proposition 10 applied to VI($\mathcal{X}$,∇f). □

Note that the only assumption of Corollary 1 which is not satisfied by the VI of Example 1 is the fact that F is a gradient operator.

The proof of equivalence between cocoercitivity and Lipschitzianity for a generic monotone gradient operator (Proposition 11) is involved and can be found in [BC10, Theorem 18.15]. Based on Figure 2.3, we provide in the following an intuition of why a monotone Lipschitz gradient operator F is cocoercive.

Let us reason for the sake of contradiction. Because F is monotone and Lipschitz, F not being cocoercive corresponds in Figure 2.3 to the image of $(1, 0)$ belonging to the x_2-axis without the origin, i.e., $F(1, 0) = (0, p)$ for some $p \neq 0$. This is the only way $F(1, 0)$ is not contained in any of the cocoercive circles (parametrized by η). Then by the multivariate mean value theorem[3] [Wik17]

$$\begin{aligned} &\left(\int_0^1 \nabla_x F\left(\begin{bmatrix}0\\0\end{bmatrix} + t\begin{bmatrix}1\\0\end{bmatrix}\right) dt\right)\left(\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}0\\0\end{bmatrix}\right) = F(1,0) - F(0,0) = \begin{bmatrix}0\\p\end{bmatrix} \\ &\Rightarrow \int_0^1 \nabla_x F(t,0) dt = \begin{bmatrix}0 & p\\ p & q\end{bmatrix} \end{aligned} \tag{2.17}$$

for some scalar q, with the integral in (2.17) meant component-wise[4]. In the implication of (2.17) we used the symmetry of the integral matrix, which is due to Proposition 2 and the fact that F is a gradient operator. Since $[\nabla_x F(t, 0)]_{(1,1)}$ must be non-negative for all $t \in [0, 1]$ (otherwise it would not be positive definite and F would not be monotone by Proposition 8), then by (2.17) $[\nabla_x F(t, 0)]_{(1,1)} = 0$ for all $t \in [0, 1]$. By (2.17) and $p \neq 0$, there must exist $\hat{t} \in [0, 1]$ such that $[\nabla_x F(\hat{t}, 0)]_{(1,2)} = [\nabla_x F(\hat{t}, 0)]_{(2,1)} \neq 0$, hence $[\nabla_x F(\hat{t}, 0)]$ is indefinite and F is not monotone by Proposition 8), which is a contradiction.

This gives the visual idea of why in the 2-dimensional case a monotone gradient operator is always cocoercive. Obviously there exists a rigorous proof for the generic n-dimensional case, but the scope here is to give an example of how Figures 2.3 and 2.5 can provide simple intuitions of implications that can be then proved rigorously.

The extragradient algorithm

Algorithm 1 finds a VI solution in presence of a strongly monotone or cocoercive operator, but its convergence is not guaranteed if the operator is only monotone. A slight variation of the projection algorithm results in the extragradient algorithm, which is guaranteed

[3] A slightly different version can be found in [Cla90, Proposition 2.6.5].

[4] The mean-value theorem requires F to be continuously differentiable, which is not needed for Proposition 11. Indeed this is just a visual explanation, Proposition 11 does not even require existence of a fixed point of F, which is instead hidden in all the derivations based on Figure 2.3.

to converge for any monotone and Lipschitz operator.

Algorithm 2: Extragradient algorithm

Initialization $\tau > 0,\ k = 0,\ x_{(0)}, \tilde{x}_{(0)} \in S$

Iterate
$$\tilde{x}_{(k+1)} = \operatorname*{Proj}_{\mathcal{X}}[x_{(k)} - \tau F(x_{(k)})]$$
$$x_{(k+1)} = \operatorname*{Proj}_{\mathcal{X}}[x_{(k)} - \tau\, F(\tilde{x}_{(k)})]$$
$$k \leftarrow k+1$$

Proposition 12 ([FP03, Theorem 12.1.11])**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be closed, convex, let $F : \mathcal{X} \to \mathbb{R}^n$ be monotone, L-Lipschitz and let VI($\mathcal{X}$,F) admit a solution. Then for any $x_{(0)}$, the sequence $(x_{(k)})_{k=0}^{\infty}$ generated by Algorithm 2 with $\tau < 1/L$ converges to a solution of VI($\mathcal{X}$,F). □

Example 1 (continued). Similarly to what done in (2.16) for Algorithm 1, we derive the expression of the update rule of Algorithm 2 applied to the linear operator in (2.15):

$$x_{(k+1)} = \operatorname*{Proj}_{\mathbb{R}^2}[x_{(k)} - \tau F(\operatorname*{Proj}_{\mathbb{R}^2}[x_{(k)} - \tau F(x_{(k)})])] = [x_{(k)} - \tau F([x_{(k)} - \tau F(x_{(k)})])] =$$
$$= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \tau \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \tau^2 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^2 \right) x_{(k)} = \begin{bmatrix} 1-\tau^2 & +\tau \\ -\tau & 1-\tau^2 \end{bmatrix} x_{(k)}. \tag{2.18}$$

The eigenvalues of the matrix in (2.18) are $1 - \tau^2 \pm i\tau$, whose square magnitude is $1 - 2\tau^2 + \tau^4 + \tau^2 = 1 - \tau^2 + \tau^4$, which is smaller than 1 for any $0 < \tau < 1$. We can conclude that for any initial condition Algorithm 2 converges to the unique solution of VI($\mathbb{R}^2$,F), whereas Example 1 shows that Algorithm 1 fails to converge. Note that the bound $\tau < 1$ corresponds to the bound of Proposition 12, because the Lipschitz constant of a linear operator is the maximum singular value of the corresponding matrix, which for the specific F equals 1. □

We conclude with a table that summarizes the sufficient conditions for convergence of Algorithms 1-2. Many other algorithms for solving monotone, cocoercive and strongly

Sufficient condition	Algorithm
F cocoercive	projection
F monotone, Lipschitz	extragradient
F monotone, Lipschitz, gradient	projection

Table 2.3: Sufficient conditions of Propositions 10, 12 and Corollary 1; further assumptions on existence of a solution, on $\mathcal{X}$ and on τ are not listed to focus on the key properties of F.

monotone VI are present in the literature; a comprehensive treatment can be found in [FP03, Chapters 10, 11, 12]. We present only Algorithms 1-2 because of their simple nature and because they are used in the remaining chapters.

2.4 Parametric variational inequalities

In this section we report a result of convergence for parametric variational inequalities, which is later used in Chapter 5 to propose a distributed algorithm for finding the Nash equilibrium of an aggregative game. To this end, we introduce the following definitions of distance between two sets.

Definition 7 (Kuratowski convergence of sets [SW79, eq. (2.1)])**.** A sequence of sets $(\mathcal{X}_\nu \subseteq \mathbb{R}^n)_{\nu=1}^\infty$ is said to Kuratowski converge to a set $\mathcal{X} \subseteq \mathbb{R}^n$, in symbols $\mathcal{X}_\nu \to \mathcal{X}$, if

$$\limsup \mathcal{X}_\nu \subseteq \mathcal{X} \subseteq \liminf \mathcal{X}_\nu, \tag{2.19}$$

where

$$\liminf \mathcal{X}_\nu := \{x \in \mathbb{R}^n | \exists\, (x_\nu)_{\nu=1}^\infty \text{ with } x_\nu \in \mathcal{X}_\nu \text{ such that } x = \lim_{\nu\to\infty} x_\nu\},$$
$$\limsup \mathcal{X}_\nu := \{x \in \mathbb{R}^n | \exists\, (\nu_k)_{k=1}^\infty, (x_{\nu_k})_{k=1}^\infty \text{ with } x_{\nu_k} \in \mathcal{X}_{\nu_k} \text{ such that } x = \lim_{k\to\infty} x_{\nu_k}\}. \qquad \square$$

By definition $\liminf \mathcal{X}_\nu \subseteq \limsup \mathcal{X}_\nu$. Condition (2.19) requires the opposite inclusion to hold.

Definition 8 (Hausdorff convergence of sets [SW79, p. 22])**.** The Hausdorff distance between two non-empty subsets $\mathcal{X}$ and $\mathcal{S}$ of $\mathbb{R}^n$ is defined as

$$d_H(\mathcal{X}, \mathcal{S}) := \max\{\sup_{s\in\mathcal{S}} \inf_{x\in\mathcal{X}} \|x - s\|_2, \sup_{x\in\mathcal{X}} \inf_{s\in\mathcal{S}} \|x - s\|_2\}\,.$$

A sequence of sets $(\mathcal{X}_\nu \subseteq \mathbb{R}^n)_{\nu=1}^\infty$ is said to Hausdorff converge to $\mathcal{X} \subseteq \mathbb{R}^n$ if

$$\lim_{\nu\to\infty} d_H(\mathcal{X}_\nu, \mathcal{X}) = 0. \qquad \square$$

It is easy to verify that the sequence $\mathcal{X}_\nu := \{(x, y) \in \mathbb{R}^2 | y \leq \nu x\}$ Kuratowski converges to $\mathcal{X} := \{(x, y) \in \mathbb{R}^2 | x \geq 0\}$, but it does not Hausdorff converge. In general Hausdorff convergence is stronger than Kuratowski convergence, in that the former implies the latter when $\mathcal{X}_\nu$ is closed for all ν [SW79, Theorem 2], while the opposite implication holds if $\mathcal{X}_\nu$ is closed for all ν and $\mathcal{X}$ is compact [SW79, Theorem 3].

We are now ready to present a classical result in convergence of parametric variational inequalities, introduced in the work [Mos69] of Umberto Mosco in 1969. In the following Proposition 13 we report a simplified version of [Mos69, Theorem A], where for the sake of readability we introduce slightly stronger assumptions compared to the original ones of [Mos69, Theorem A].

Proposition 13. Let us consider the closed, convex set $\mathcal{X} \subseteq \mathbb{R}^n$ and the operator $F : \mathcal{X} \to \mathbb{R}^n$. Let us consider the sequence of closed, convex sets $(\mathcal{X}_\nu)_{\nu=1}^\infty$ and the sequence of operators $(F_\nu)_{\nu=1}^\infty$, with $F_\nu : \mathcal{X}_\nu \to \mathbb{R}^n$. Assume that

1. F is continuous and monotone, F_ν is continuous and monotone for all ν, and $\lim_{\nu\to\infty} F_\nu(x) = F(x)$ for all $x \in \mathcal{X}$;
2. VI($\mathcal{X}$,F) admits a unique solution $\bar{x}$;
3. $\mathcal{X}_\nu$ is bounded uniformly in ν;
4. $\mathcal{X}_\nu$ Kuratowski converges to $\mathcal{X}$.

Then the sequence $(x_\nu)_{\nu=1}^\infty$, where x_ν is a solution of VI($\mathcal{X}_\nu$,F_ν), converges to $\bar{x}$. □

CHAPTER 3

Nash and Wardrop equilibria in aggregative games

3.1 Game and equilibria

Motivated by the considerations of Chapter 1, we consider a set of M agents. Each agent can choose his strategy x^i in his individual constraint set $\mathcal{X}^i \subset \mathbb{R}^n$. For finite horizon discrete-time dynamic games the constraint set can represent the dynamics of a system [MCH13, GPC16], as in the electric vehicle application of Section 6.1. We assume that the cost function

$$J^i(x^i, \sigma(x)) \tag{3.1}$$

of agent i depends on his own strategy $x^i \in \mathcal{X}^i$ and on the strategies of the other agents via the average strategy $\sigma(x) := \frac{1}{M}\sum_{j=1}^M x^j \in \frac{1}{M}\sum_{j=1}^M \mathcal{X}^j$, as typical of aggregative games [Jen10]. Besides the individual constraints, each agent has to satisfy a coupling constraint, which involves the decision variables of other agents. Upon defining the vector of stacked-strategies as $x = [x^1; \ldots; x^M] \in \mathbb{R}^{Mn}$, the coupling constraint can be expressed as

$$x \in \mathcal{C} := \{x \in \mathbb{R}^{Mn} \,|\, g(x) \le \mathbb{0}_m\} \subset \mathbb{R}^{Mn}, \tag{3.2}$$

with $g : \mathbb{R}^{Mn} \to \mathbb{R}^m$. The coupling constraint in (3.2) can model for instance the fact that the overall usage level for a certain commodity cannot exceed a fixed capacity, as in the applications of Chapter 6. The cost and constraints just introduced give rise to the game

$$\mathcal{G} := \begin{cases} \text{agents :} & \{1, \ldots, M\} \\ \text{cost of agent } i \text{ :} & J^i(x^i, \sigma(x)) \\ \text{individual constraint :} & \mathcal{X}^i \\ \text{coupling constraint :} & \mathcal{C}, \end{cases} \tag{3.3}$$

which is the focus of Chapters 3, 4, 5 and 6. We denote for convenience $\mathcal{X} := \mathcal{X}^1 \times \ldots \times \mathcal{X}^M$ and define

$$\mathcal{Q}^i(x^{-i}) := \{x^i \in \mathcal{X}^i \,|\, g(x) \le \mathbb{0}_m\}, \qquad \mathcal{Q} := \mathcal{X} \cap \mathcal{C}. \tag{3.4}$$

If there is no coupling constraint, i.e., $\mathcal{C} = \mathbb{R}^{Mn}$, then $\mathcal{G}$ becomes a game described in normal form as in [Aub07, Definition 1]. We point out that the entire thesis focuses on continuous strategy spaces [Aub07] rather than discrete strategy spaces [NRT07]. The reader interested in a specific instance of the game $\mathcal{G}$ in (3.3) can read the first two pages of one of the applications of Chapter 6.

3.1.1 Equilibrium definitions

We consider two notions of equilibrium for the game $\mathcal{G}$ in (3.3), namely the Nash and the Wardrop equilibrium. The concept of Nash equilibrium was originally formulated for games without coupling constraints in the seminal works of Von Neumann [VNM45] and Nash [Nas51], and then extended to games with coupling constraints by Debreu [AD54] and Rosen [Ros65].

Definition 9 (Nash Equilibrium)**.** A vector of strategies $x_{\mathrm{N}} = [x_{\mathrm{N}}^1; \dots ; x_{\mathrm{N}}^M] \in \mathbb{R}^{Mn}$ is an ε-Nash equilibrium of the game $\mathcal{G}$ if $x_{\mathrm{N}} \in \mathcal{Q}$ and for all $i \in \{1, \dots , M\}$ and all $x^i \in \mathcal{Q}^i(x_{\mathrm{N}}^{-i})$

$$J^i(x_{\mathrm{N}}^i, \sigma(x_{\mathrm{N}})) \leq J^i\left(x^i, \tfrac{1}{M}x^i + \tfrac{1}{M}\textstyle\sum_{j\neq i} x_{\mathrm{N}}^j\right) + \varepsilon\,. \tag{3.5}$$

If (3.5) holds with $\varepsilon = 0$ then x_{N} is a Nash equilibrium. □

Intuitively, a feasible set of strategies $\{x_{\mathrm{N}}^i\}_{i=1}^M$ is a Nash equilibrium if no agent can improve his cost by unilaterally deviating from his strategy, assuming that the strategies of the other agents are fixed. When the game features no coupling constraints, i.e., $\mathcal{C} = \mathbb{R}^n$, then in Definition 9 the unilateral strategy change x^i is constrained to be in $\mathcal{X}^i$ rather than in $\mathcal{Q}^i(x_{\mathrm{N}}^{-i})$. A Nash equilibrium for a game with coupling constraints is usually referred to as generalized Nash equilibrium [FK07]; in this thesis we omit the word generalized for brevity, even though we consider a game with coupling constraints.

Note that on the right-hand side of (3.5) the decision variable x^i appears in both arguments of $J^i(\cdot, \cdot)$. However, as the number of agents grows the contribution of agent i to the average $\sigma(x)$ decreases. This observation motivates the definition of Wardrop equilibrium.

Definition 10 (Wardrop Equilibrium)**.** A vector of- strategies $x_{\mathrm{W}} = [x_{\mathrm{W}}^1; \dots ; x_{\mathrm{W}}^M] \in \mathbb{R}^{Mn}$ is a Wardrop equilibrium of the game $\mathcal{G}$ if $x_{\mathrm{W}} \in \mathcal{Q}$ and for all $i \in \{1, \dots , M\}$ and all $x^i \in \mathcal{Q}^i(x_{\mathrm{W}}^{-i})$

$$J^i(x_{\mathrm{W}}^i, \sigma(x_{\mathrm{W}})) \leq J^i(x^i, \sigma(x_{\mathrm{W}})).$$ □

Intuitively, a feasible set of strategies $\{x_{\mathrm{W}}^i\}_{i=1}^M$ is a Wardrop equilibrium if no agent can improve his cost by unilaterally deviating from his strategy, assuming that the average strategy is fixed.

The next examples clarifies the origin of the term "Wardrop equilibrium" by showing that Definition 10 generalizes to generic aggregative games the notion of Wardrop equilibrium introduced in [War52, Smi79, Daf80] for traffic networks.

Example 2 (Wardrop equilibrium for parallel roads)**.** We consider a game where each agent $i \in \{1, 2, ..., M\}$ wants to send $\gamma^i \in \mathbb{R}_{>0}$ units of mass from origin to destination, which are connected by n parallel roads as in Figure 3.1. In this simple example, all the agents share the same origin and destination. The strategy x^i represents the distribution of mass γ^i across the roads, hence the constraint set of agent i is the simplex

$$x^i \in \mathcal{X}^i := \left\{x^i \in \mathbb{R}^n_{\geq 0} | \mathbb{1}_n^\top x^i = \gamma^i \right\}.$$

Each road $j \in \{1, 2, ..., n\}$ is associated with a travel time $t_j(\sigma(x))$ which is a function of the average mass distribution across the roads. The cost function of agent i represents his total travel time:

$$J(x^i, \sigma(x)) = \sum_{j=1}^{n} t_j(\sigma(x)) x^i_j = t(\sigma(x))^\top x^i,$$

where we denoted $t(\sigma(x)) = [t_j(\sigma(x))]_{j=1}^n$. This game features no coupling constraints, i.e., $\mathcal{C} = \mathbb{R}^n$.

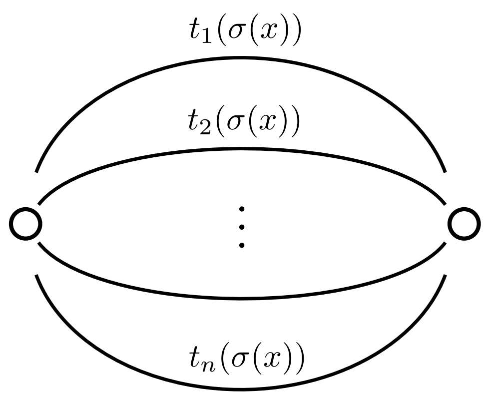


Figure 3.1: A parallel road network with with the travel times of the different roads.

Let us consider a Wardrop equilibrium x_{W} and equivalently rewrite the no-improvement condition of Definition 10 in terms of $\sigma(x_\mathrm{W})$:

$$\begin{aligned}
&x_\mathrm{W} \text{ is Wardrop equilibrium } \Leftrightarrow t(\sigma(x_\mathrm{W}))^\top x_\mathrm{W}^i \leq t(\sigma(x_\mathrm{W}))^\top x^i, \quad \forall\, x^i \in \mathcal{X}^i,\ \forall i \\
&\overset{(\star)}{\Leftrightarrow} \frac{1}{M}\sum_{i=1}^M t(\sigma(x_\mathrm{W}))^\top x_\mathrm{W}^i \leq \frac{1}{M}\sum_{i=1}^M t(\sigma(x_\mathrm{W}))^\top x^i, \quad \forall\, x \in \mathcal{X} \\
&\Leftrightarrow t(\sigma(x_\mathrm{W}))^\top \sigma(x_\mathrm{W}) \leq t(\sigma(x_\mathrm{W}))^\top \sigma(x), \quad \forall\, x \in \mathcal{X} \\
&\Leftrightarrow t(\sigma(x_\mathrm{W}))^\top \sigma(x_\mathrm{W}) \leq t(\sigma(x_\mathrm{W}))^\top \sigma, \quad \forall\, \sigma \in \frac{1}{M}\sum_{i=1}^M \mathcal{X}^i. \\
&\Leftrightarrow \sigma(x_\mathrm{W}) \in \underset{\sigma \in \frac{1}{M}\sum_{i=1}^M \mathcal{X}^i}{\operatorname{argmin}} \; t(\sigma(x_\mathrm{W}))^\top \sigma. \qquad (3.6)
\end{aligned}$$

The equivalence $(\star)$ is straightforward in the direction $\Rightarrow$, whereas the direction $\Leftarrow$ for a specific i can be seen by fixing $x^{-i} = x_\mathrm{W}^{-i}$, as all the summands vanish except the i^th one. Upon defining $\gamma_\text{avg} := \frac{1}{M}\sum_{i=1}^M \gamma^i$, one can rewrite

$$\frac{1}{M}\sum_{i=1}^M \mathcal{X}^i = \left\{\sigma \in \mathbb{R}^E_{\geq 0} | \mathbb{1}_n^\top \sigma = \gamma_\text{avg}\right\}.$$

Defining the Wardrop equilibrium in terms of the aggregate quantity $\sigma(x_\mathrm{W})$ as in (3.6), rather than the individual strategies x_W, is standard in all the literature of transportation engineering. Such aggregate equilibrium is also referred to as traffic user equilibrium [Daf80] and it is defined and analyzed not only for a parallel road network, but for a generic network with different origin-destination pairs [Smi79], as we study in Section 6.2.

In the following, we show how (3.6) corresponds to the original definition of equilibrium [War52] given in words by Wardrop in 1952. This correspondence has already been derived in [Daf80], but we report it nonetheless for its instructive nature. By Proposition 3, $\sigma(x_\mathrm{W})$ is a solution of the optimization program in (3.6) if and only if it solves its KKT system:

$$\begin{aligned}
&t(\sigma(x_\mathrm{W})) - \lambda + \mu \mathbb{1}_n = \mathbb{0}_n && (3.7a)\\
&\mathbb{0}_n \leq \lambda \perp \sigma(x_\mathrm{W}) \geq \mathbb{0}_n, && (3.7b)\\
&\mathbb{1}_n^\top \sigma(x_\mathrm{W}) = \gamma_\text{avg}, && (3.7c)
\end{aligned}$$

where $\lambda \in \mathbb{R}^n$ is the dual variable corresponding to the non-negativity constraint and $\mu \in \mathbb{R}$ is the dual variable corresponding to the constraint $\mathbb{1}_n^\top \sigma = \gamma$. By substituting (3.7a) into (3.7b), the system (3.7) can be simplified into

$$\begin{aligned}
&\mathbb{0}_n \leq t(\sigma(x_\mathrm{W})) + \mu \mathbb{1}_n \perp \sigma(x_\mathrm{W}) \geq \mathbb{0}_n, && (3.8a)\\
&\mathbb{1}_n^\top \sigma(x_\mathrm{W}) = \gamma_\text{avg}. && (3.8b)
\end{aligned}$$

We now argue that it must hold

$$\mu = -t_\text{min}(\sigma(x_\mathrm{W})) := -\min_{j \in \{1,\dots,n\}} t_j(\sigma(x_\mathrm{W})).$$

Indeed, if $\mu > -t_{\min}(\sigma(x_{\mathrm{W}}))$ then $t(\sigma(x_{\mathrm{W}})) + \mu\mathbb{1}_n > \mathbb{0}_n$, hence by the orthogonality condition (3.8a) $\sigma(x_{\mathrm{W}}) = \mathbb{0}_n$ which violates (3.8b). If instead $\mu < -t_{\min}(\sigma(x_{\mathrm{W}}))$ then $t(\sigma(x_{\mathrm{W}})) + \mu\mathbb{1}_n \geq \mathbb{0}_n$ does not hold. We can conclude that (3.8) reads as

$$\begin{aligned}&\mathbb{0}_n \leq t(\sigma(x_{\mathrm{W}})) - t_{\min}(\sigma(x_{\mathrm{W}}))\mathbb{1}_n \perp \sigma(x_{\mathrm{W}}) \geq \mathbb{0}_n,\\ &\mathbb{1}_n^\top \sigma(x_{\mathrm{W}}) = \gamma_{\mathrm{avg}},\end{aligned}$$

which is equivalent to $\sigma(x_{\mathrm{W}}) \geq \mathbb{0}_n, \mathbb{1}_n^\top \sigma(x_{\mathrm{W}}) = \gamma_{\mathrm{avg}}$ and

$$\begin{aligned}\sigma_j(x_{\mathrm{W}}) > 0 &\Rightarrow t_j(\sigma(x_{\mathrm{W}})) = t_{\min}(\sigma(x_{\mathrm{W}}))\\ \sigma_j(x_{\mathrm{W}}) = 0 &\Rightarrow t_j(\sigma(x_{\mathrm{W}})) \geq t_{\min}(\sigma(x_{\mathrm{W}})).\end{aligned} \tag{3.10}$$

In words, "all the used roads feature minimum travel time and there is no unused road more convenient than a used one". This is indeed the celebrated Wardrop user equilibrium principle [War52], which is at the core of contemporary transportation engineering, for stationary traffic networks [CSM11, She85] as well as for dynamic ones [Smi79, FBS93]. □

Comparison with the literature

Even though the Wardrop equilibrium is a classical concept, the existing literature on aggregative games [ABS02b, AW04, ABEA06, ACA11, MW95, DN87] focuses on the aggregate formulation (3.10), defining the Wardrop equilibrium in terms of $\sigma(x)$, whereas Definition 10 is expressed in terms of the agents' strategies x. The first glimmer of Wardrop equilibrium in terms of x appears in the works [MCH13, GPC16], where however it is not recognized as an equilibrium concept on its own and connected to the classical Wardrop in terms of $\sigma(x)$, but rather only identified as an ε-Nash. To the best of our knowledge, ours is the first effort to define a Wardrop equilibrium in terms of x. This appears to be a natural attempt, as it is common understanding that Nash and Wardrop equilibria are closely related, but to better characterize the distance between the two, they need to be defined on the same space, namely the one of the agents' strategies $x \in \mathbb{R}^{Mn}$. As a bibliographical note, we point out that the work [Gra17], which appeared online a couple of days before our manuscript [GPP17], also defines the Wardrop equilibrium in terms of x by naming it *aggregative equilibrium*, but fails to recognize its connection with the classic Wardrop equilibrium expressed in terms of $\sigma(x)$.

It is also worth mentioning that we define such concept for a generic aggregative game, with arbitrary individual constraint sets $\mathcal{X}^i$ and cost function $J^i(x^i, \sigma(x))$, whereas the aforementioned works [ABS02b, AW04, ABEA06, ACA11, MW95, DN87] focus on specific applications, as transportation networks, competitive markets or charging of electric vehicles, and hence each of them considers either specific constraint sets or specific cost functions. Indeed for generic constraint sets $\mathcal{X}^i$ and cost function $J^i(x^i, \sigma(x))$ it is not possible to express the Wardrop equilibrium in terms of $\sigma(x)$, as we instead did in Example 2, specifically in reformulation (3.6). So studying the Wardrop equilibrium in terms of x is not a mere change of perspective, but rather a new approach that can address a larger class of equilibrium problems.

3.2 Relation with variational inequalities

The results of Sections 3.4 and of Chapters 4 and 5 are based on the fact that certain Nash and Wardrop equilibria of the game $\mathcal{G}$ in (3.3) can be obtained by solving a variational inequality, which is the subject of this section. This is a well-known result which follows from the first-order optimality condition of Proposition 1. For finite dimensional spaces, such result was shown for games without coupling constraints in [Ben74, eq. (1.11)] and for games with coupling constraints in [Har91, Theorem 3], then generalized in [FFP07, Theorem 2.1]. This result allows one to employ the powerful mathematical tool of variational inequalities (see Chapter 2) to study and find equilibria of the game. Let us define

$$F_{\mathrm{N}}(x) \coloneqq [\nabla_{x^i} J^i(x^i, \sigma(x))]_{i=1}^M = \nabla_{z_1} J^i(x^i, \sigma(x)) + \frac{1}{M} \nabla_{z_2} J^i(x^i, \sigma(x)), \quad (3.11a)$$

$$F_{\mathrm{W}}(x) \coloneqq [\nabla_{x^i} J^i(x^i, z)_{|z=\sigma(x)}]_{i=1}^M = \nabla_{z_1} J^i(x^i, \sigma(x)), \quad (3.11b)$$

where $F_{\mathrm{N}},\ F_{\mathrm{W}} : \mathcal{X} \to \mathbb{R}^{Mn}$ and where we used the notation

$$\nabla_{z_1} J^i(x^i, \sigma(x)) = \nabla_{z_1} J^i(z_1, z_2)_{|z_1=x^i, z_2=\sigma(x)},$$
$$\nabla_{z_2} J^i(x^i, \sigma(x)) = \nabla_{z_2} J^i(z_1, z_2)_{|z_1=x^i, z_2=\sigma(x)}.$$

The operator F_{N} is obtained by stacking together the gradients of each agent's cost with respect to his decision variable. F_{W} is obtained similarly, but considering $\sigma(x)$ as fixed when differentiating. The following proposition provides a sufficient characterization of the equilibria described in Definitions 9 and 10 as solutions of two variational inequalities, which feature the same set $\mathcal{Q}$, defined in (3.4), but different operators, namely F_{N} and F_{W} in (3.11).

Assumption 1. For all $i \in \{1, \dots, M\}$, the individual constraint set $\mathcal{X}^i$ is closed and convex. The set $\mathcal{Q}$ in (3.4) has non-empty interior. The cost functions $J^i(x^i, \sigma(x))$ are convex in x^i for any fixed $\{x^j \in \mathcal{X}^j\}_{j \neq i}$. The cost functions $J^i(x^i, z)$ are convex in x^i for any $z \in \frac{1}{M}\sum_{j=1}^M \mathcal{X}^j$. The cost functions $J^i(z_1, z_2)$ are continuously differentiable in $[z_1; z_2]$ for any $z_1 \in \mathcal{X}^i$ and $z_2 \in \frac{1}{M}\sum_{j=1}^M \mathcal{X}^j$. The function g in (3.2) is convex. □

Proposition 14. Under Assumption 1, the following hold.

1. Any solution $\bar{x}_{\mathrm{N}}$ of $\mathrm{VI}(\mathcal{Q}, F_{\mathrm{N}})$ is a Nash equilibrium of the game $\mathcal{G}$ in (3.3);

2. Any solution $\bar{x}_{\mathrm{W}}$ of $\mathrm{VI}(\mathcal{Q}, F_{\mathrm{W}})$ is a Wardrop equilibrium of the game $\mathcal{G}$ in (3.3). □

Proof (based on first-order condition). The proof of the first statement can be found in [FFP07, Theorem 2.1], but we report it here as it is fundamental and instructional. By definition, if $\bar{x}_{\mathrm{N}}$ solves $\mathrm{VI}(\mathcal{Q}, F_{\mathrm{N}})$ then

$$F_{\mathrm{N}}(\bar{x}_{\mathrm{N}})^\top (x - \bar{x}_{\mathrm{N}}) \geq 0, \quad \forall\, x \in \mathcal{Q}. \tag{3.12}$$

Consider $i \in \{1, \dots, M\}$, set $x^{-i} = \bar{x}_{\mathrm{N}}^{-i}$ in (3.12) and consider an arbitrary $x^i \in \mathcal{Q}^i(\bar{x}_{\mathrm{N}}^{-i})$; then all the summands in (3.12) vanish except the i^{th} one and (3.12) reads

$$\nabla_{x^i} J^i(\bar{x}_{\mathrm{N}}^i, \sigma(\bar{x}_{\mathrm{N}}))^\top (x^i - \bar{x}_{\mathrm{N}}^i) \geq 0,\ \forall\, x^i \in \mathcal{Q}^i(\bar{x}_{\mathrm{N}}^{-i}). \tag{3.13}$$

Consider the convex function $x^i \to J^i(x^i, \frac{1}{M} x^i + \frac{1}{M} \sum_{j \neq i} \bar{x}_{\mathrm{N}}^j) : \mathcal{Q}^i(\bar{x}_{\mathrm{N}}^{-i}) \to \mathbb{R}$. Since $\mathcal{Q}^i(\bar{x}_{\mathrm{N}}^{-i})$ is a convex set, the first-order condition (3.13) is equivalent to optimality of $\bar{x}_{\mathrm{N}}^i$ by Proposition 1, that is,

$$(3.13) \Leftrightarrow J^i\left(\bar{x}_{\mathrm{N}}^i, \sigma(\bar{x}_{\mathrm{N}})\right) \leq J^i\left(x^i, \frac{1}{M} x^i + \frac{1}{M} \sum_{j \neq i} \bar{x}_{\mathrm{N}}^j\right), \quad \forall\, x^i \in \mathcal{Q}^i(\bar{x}_{\mathrm{N}}^{-i}).$$

As this holds for all $i \in \{1, \dots, M\}$ and since $\bar{x}_{\mathrm{N}} \in \mathcal{Q}$, then $\bar{x}_{\mathrm{N}}$ is a Nash equilibrium of $\mathcal{G}$ by Definition 9.

The proof of the second statement is analogous to the first one, but we report it nonetheless for its novelty. We rewrite the operator $F_{\mathrm{W}}(x)$ as $\tilde{F}_{\mathrm{W}}(x, \sigma(x))$, where $\tilde{F}_{\mathrm{W}}(x, z) := [\nabla_{x^i} J^i(x^i, z)]_{i=1}^M$. By definition, if $\bar{x}_{\mathrm{W}}$ solves $\mathrm{VI}(\mathcal{Q}, F_{\mathrm{W}})$ then $F_{\mathrm{W}}(\bar{x}_{\mathrm{W}})^\top (x - \bar{x}_{\mathrm{W}}) \geq 0$ for all $x \in \mathcal{Q}$, i.e.

$$\tilde{F}_{\mathrm{W}}(\bar{x}_{\mathrm{W}}, \bar{z}_{\mathrm{W}})^\top (x - \bar{x}_{\mathrm{W}}) \geq 0,\ \forall x \in \mathcal{Q}, \tag{3.14}$$

where $\bar{z}_{\mathrm{W}} = \sigma(\bar{x}_{\mathrm{W}})$. Consider $i \in \{1, \dots, M\}$, set $x^{-i} = \bar{x}_{\mathrm{W}}^{-i}$ in (3.14) and consider an arbitrary $x^i \in \mathcal{Q}^i(\bar{x}_{\mathrm{W}}^{-i})$; then all the summands in (3.14) vanish except the i^{th} one and (3.14) reads

$$\nabla_{x^i} J^i(\bar{x}_{\mathrm{W}}^i, \bar{z}_{\mathrm{W}})^\top (x^i - \bar{x}_{\mathrm{W}}^i) \geq 0,\ \forall\, x^i \in \mathcal{Q}^i(\bar{x}_{\mathrm{W}}^{-i}). \tag{3.15}$$

Consider the convex function $x^i \to J^i(x^i, \bar{z}_{\mathrm{W}}) : \mathcal{Q}^i(\bar{x}_{\mathrm{W}}^{-i}) \to \mathbb{R}$. Since $\mathcal{Q}^i(\bar{x}_{\mathrm{W}}^{-i})$ is a convex set, the first-order condition (3.15) is equivalent to optimality of $\bar{x}_{\mathrm{W}}$ by Proposition 1, that is,

$$(3.15) \Leftrightarrow \bar{x}_{\mathrm{W}}^i \in \operatorname*{argmin}_{x^i \in \mathcal{Q}^i(\bar{x}_{\mathrm{W}}^{-i})} J^i\left(x^i, \bar{z}_{\mathrm{W}}\right).$$

Substituting $\bar{z}_{\mathrm{W}} = \sigma(\bar{x}_{\mathrm{W}})$, one has $J^i\left(\bar{x}_{\mathrm{W}}^i, \sigma(\bar{x}_{\mathrm{W}})\right) \leq J^i\left(x^i, \sigma(\bar{x}_{\mathrm{W}})\right)$ for all $x^i \in \mathcal{Q}^i(\bar{x}_{\mathrm{W}}^{-i})$. As this holds for all $i \in \{1, \dots, M\}$ and since $\bar{x}_{\mathrm{W}} \in \mathcal{Q}$, then $\bar{x}_{\mathrm{W}}$ is a Wardrop equilibrium of $\mathcal{G}$ by Definition 10. □

Alternative proof (based on KKT system). We conduct an alternative proof of Proposition 14 based on the KKT system, which requires the additional Slater's constraint qualification but gives a further insight into the relation between equilibria and VI solutions in terms of the dual variables relative to the coupling constraints. This alternative proof is well known in the literature and can be found in [FK07, Theorem 8]. We report only the one relative to the Nash equilibrium, the one for Wardrop being analogous.

For ease of readability we assume that $\mathcal{X}^i = \mathbb{R}^n$ for all i, but this can be easily generalized to any closed and convex set $\mathcal{X}^i$, see [FK07, Theorem 9]. Further, we must assume that $\mathcal{Q} = \{x \in \mathbb{R}^{Mn} | \, g(x) \leq \mathbb{0}_m\}$ in (3.4) satisfies Slater's constraint qualification of Definition 2 Then by Proposition 4 VI$(\mathcal{Q}, F_\mathrm{N})$ is equivalent to its KKT system

$$\begin{aligned} &F_\mathrm{N}(x) + \nabla_x g(x)\lambda = \mathbb{0}_{Mn} \\ &\mathbb{0}_m \leq \lambda \perp g(x) \leq \mathbb{0}_m \end{aligned}$$

which is equivalent to

$$\nabla_{x^i} J^i(x, \sigma(x)) + \nabla_{x^i} g(x)\lambda = \mathbb{0}_n \qquad \forall\, i \in \{1, \dots, M\} \tag{3.17a}$$

$$\mathbb{0}_m \leq \lambda \perp g(x) \leq \mathbb{0}_m \qquad \forall\, i \in \{1, \dots, M\}. \tag{3.17b}$$

Condition (3.17b) is redundantly repeated M times to draw a sharper parallel with the following KKT system (3.18). By Proposition 3, the optimality condition (3.5) for the Nash equilibrium is equivalent to M KKT systems, one for each agent:

$$\nabla_{x^i} J^i(x, \sigma(x)) + \nabla_{x^i} g(x)\lambda^i = \mathbb{0}_n \qquad \forall i \in \{1, \dots, M\} \tag{3.18a}$$

$$\mathbb{0}_m \leq \lambda^i \perp g(x) \leq \mathbb{0}_m \qquad \forall i \in \{1, \dots, M\}. \tag{3.18b}$$

The difference between (3.17) and (3.18) is that in the former each agent must share the same dual variable $\lambda \in \mathbb{R}^n$, while in the latter each agent can choose his own $\lambda^i \in \mathbb{R}^n$. It is then clear that any solution $\bar{x}_\mathrm{N}$ to (3.17) (i.e., $\bar{x}_\mathrm{N}$ such that there exists $\bar{\lambda}$ with $(\bar{x}_\mathrm{N}, \bar{\lambda})$ solving (3.17)) is also a solution to (3.18) (i.e., $\bar{x}_\mathrm{N}$ such that there exists $\{\bar{\lambda}^i\}_{i=1}^M$ with $(\bar{x}_\mathrm{N}, \{\bar{\lambda}^i\}_{i=1}^M)$ solving (3.18)), but the viceversa is not true in general. □

Proposition 14 states that a solution of the variational inequality is an equilibrium. The converse in general does not hold, due to the presence of the coupling constraints, as highlighted in the alternative proof of Proposition 14. This can also be seen in the first proof, because (3.13) does not imply (3.12). The equilibria that can be obtained as solutions of the corresponding variational inequality are called *variational equilibria* [FK07, Definition 3] and in this Chapter 3, as well as in Chapters 4 and 5 they are denoted with $\bar{x}_\mathrm{N}, \bar{x}_\mathrm{W}$ instead of $x_\mathrm{N}, x_\mathrm{W}$, that indicate any equilibria satisfying Definitions 9 and 10.

If on the other hand $\mathcal{C} = \mathbb{R}^{Mn}$, then $\mathcal{Q} = \mathcal{X}^1 \times \dots \mathcal{X}^M$ and $\mathcal{Q}^i(\bar{x}_\mathrm{N}^{-i}) = \mathcal{X}^i$, thus (3.13) implies (3.12). This fact is even more direct to see in the alternative proof, where the dual variable relative to the coupling constraint does not exist in the first place. As a consequence, for a game without coupling constraints x_N solves the VI$(\mathcal{Q}, F_\mathrm{N})$ if and only if it is a Nash equilibrium of $\mathcal{G}$ and x_W solves the VI$(\mathcal{Q}, F_\mathrm{W})$ if and only if it is a Wardrop equilibrium of $\mathcal{G}$ [FK07, Corollary 1]. In other words, variational inequalities completely characterize Nash and Wardrop equilibria of a convex game without coupling constraint. The survey [FK07] presents an instructive example of a simple game with coupling constraints admitting multiple Nash equilibria but only one variational equilibrium (Example 1, p. 175 and its continuation, p. 186).

Variational and normalized equilibria

The work by Rosen [Ros65] is one of the first to study games with coupling constraints and introduces the concept of *normalized Nash equilibrium*, that is a vector x_N for which there exist weights $r \in \mathbb{R}^M_{>0}$ such that x_N solves the VI$(\mathcal{Q}, F^r_\mathrm{N})$, where $F^r_\mathrm{N}(x) \coloneqq [r_i \nabla_{x^i} J^i(x^i, \sigma(x))]_{i=1}^M$. The variational Nash equilibrium (as introduced in the previous section) is a special case of normalized Nash equilibrium, because it corresponds to $r = \mathbb{1}_M$. Following a very similar proof to that of Proposition 14, one can show that any normalized Nash equilibrium is a Nash equilibrium, so that

$$x \text{ variational equilibrium} \;\Rightarrow\; x \text{ normalized equilibrium} \;\Rightarrow\; x \text{ equilibrium.}$$

The opposite implications do not hold in general, as it is shown in Example 1 [FK07, p. 188], but in games without coupling constraints the three concepts coincide, as explained in the previous subsection.

It is proven in [Ros65] that the choice of r corresponds to a split of the burden of satisfying the coupling constraint among the agents. The intuition behind this lays in the fact that finding a normalized equilibrium amounts to solving (3.17) where the first summand $\nabla_{x^i} J^i(x, \sigma(x))$ is pre-multiplied by $r^i > 0$, which is equivalent to dividing the common dual variable λ by r^i in the second summand, so that by solving VI$(\mathcal{Q}, F^r_\mathrm{N})$ the agents are assigned different dual variables based on r. The lower r^i, the higher the responsibility for agent i of satisfying the coupling constraint.

In the context of aggregative games, however, each agent contributes equally to the average and in all the applications of Chapter 6 we study coupling constraints expressed on the average. Therefore we split the burden of the coupling constraint equally among the agents by selecting $r = \mathbb{1}_M$, that is, we focus on variational equilibria, as typically done in the aggregative game literature [FK07, PP09, FFP07]. Nonetheless we note that our results could be easily extended to normalized equilibria by using the operator F^r_N instead of F_N. The above discussion was conducted for Nash, but the same arguments are also valid for the Wardrop equilibrium.

3.3 Sufficient conditions for strong monotonicity

As variational equilibria are solutions of variational inequalities, one can use the VI results of Chapter 2 to establish existence and uniqueness of variational equilibria.

Proposition 15 ([Ros65, Theorems 1 and 2])**.** Under Assumption 1,

1. if the set $\mathcal{Q}$ is bounded then there exists a variational Nash equilibrium (Proposition 6);
2. if the operator F_{N} is strongly monotone then there exists a unique variational Nash equilibrium (Proposition 7), but there might be multiple Nash equilibria.

The same statements hold for variational Wardrop equilibrium.

Proof. The statements are trivial consequences of the VI reformulation in Proposition 14, and of the results on existence and uniqueness of VI solutions in Propositions 6 and 7. □

This section is dedicated to deriving sufficient conditions that guarantee strong monotonicity of the VI operators, which is not only needed to guarantee uniqueness of the variational equilibrium, but it is also crucial for the results of proximity of Nash and Wardrop of the next Section 3.4 and for convergence of the algorithms of Chapters 4 and 5. In particular, each of the next three subsections considers a special form of (3.1) and provides sufficient conditions for monotonicity and strong monotonicity of F_{N} and F_{W}.

Generic price function

We specialize the cost function (3.1) of agent i to

$$J^i(x^i, \sigma(x)) := v^i(x^i) + p(\sigma(x))^\top x^i. \tag{3.19}$$

The cost in (3.19) can describe applications where x^i denotes the usage level of a certain commodity, whose negative utility is modeled by $v^i : \mathcal{X}^i \to \mathbb{R}$ and whose per-unit cost $p : \frac{1}{M}\sum_{i=1}^M \mathcal{X}^i \to \mathbb{R}^n$ depends on the average usage level [CLL14, MCH13]. Chapter 6 indeed focuses on three applications which all share the form (3.19). The operators in (3.11) become

$$F_{\mathrm{W}}(x) = [\nabla_{x^i} v^i(x^i)]_{i=1}^M + [p(\sigma(x))]_{i=1}^M, \tag{3.20a}$$

$$F_{\mathrm{N}}(x) = F_{\mathrm{W}}(x) + \frac{1}{M}[\nabla_z p(z)_{|z=\sigma(x)} x^i]_{i=1}^M. \tag{3.20b}$$

The next lemma makes use of Proposition 8, together with the relation between convexity and monotonicity of Table 2.1, to derive simple sufficient conditions for strong monotonicity of the operators F_{N} and F_{W} in (3.20).

Lemma 2. The following hold.

1. Suppose that for each agent $i \in \{1, \dots, M\}$ the function v^i in (3.19) is convex and that p is monotone; then F_{W} is monotone. Under the further assumption that p is affine and strongly monotone, F_{N} is strongly monotone.

2. Suppose that for each agent $i \in \{1, \dots, M\}$ the function v^i in (3.19) is strongly convex and that p is monotone. Then F_{W} is strongly monotone. □

Proof. 1) Let us first show that F_{W} is monotone. Since v^i is convex, then $\nabla_{x^i} v^i(x^i)$ is monotone in x^i by Table 2.1. Hence $[\nabla_{x^i} v^i(x^i)]_{i=1}^M$ is monotone. Moreover, for any x_1, x_2

$$\begin{aligned}&([p(\sigma(x_1))]_{i=1}^M - [p(\sigma(x_2))]_{i=1}^M)^\top (x_1 - x_2)\\ &= M(p(\sigma(x_1)) - p(\sigma(x_2)))^\top (\sigma(x_1) - \sigma(x_2)) \geq 0,\end{aligned} \tag{3.21}$$

where the last inequality follows from the fact that p is monotone. By (3.20a) and the fact that the sum of two monotone operators is monotone by Lemma 1, one can conclude that F_{W} is monotone.
To show that F_{N} is strongly monotone, we write the affine expression of p as $p(x) = Cx + c$, where there exists $\alpha > 0$ such that $C \succ \alpha I_n$ by Proposition 8. Then the term $\frac{1}{M}[\nabla_z p(z)_{|z=\sigma(x)} x^i]_{i=1}^M$ in (3.20b) equals $\frac{1}{M}(I_M \otimes C^\top)x$. Since $\nabla_x(\frac{1}{M}(I_M \otimes C^\top)x) \succ \frac{\alpha}{M} I_{Mn}$, then $\frac{1}{M}[\nabla_z p(z)_{|z=\sigma(x)} x^i]_{i=1}^M$ is strongly monotone by Proposition 8. Having already shown that F_{W} is monotone, the proof is concluded upon noting that the sum of a monotone operator and a strongly monotone operator is strongly monotone by Lemma 1.

2) Strong convexity of v^i is equivalent to strong monotonicity of $\nabla_{x^i} v^i(x^i)$ in x^i by Table 2.1. Then $[\nabla_{x^i} v^i(x^i)]_{i=1}^M$ is strongly monotone. Monotonicity of $[p(\sigma(x))]_{i=1}^M$ in (3.20a) can be shown as in (3.21). □

Contrary to the next two subsections, we do not provide a specific literature comparison for Lemma 2, as its statement and proof are elementary.

Diagonal price function

We now focus on a cost function which is even more specific than (3.19),

$$J^i(x^i, \sigma(x)) = v^i(x^i) + \sum_{t=1}^n p_t(\sigma_t(x_t)) x_t^i =: v^i(x^i) + p(\sigma(x))^\top x^i. \tag{3.22}$$

We refer to p as *diagonal price function*, because the t^{th} component of p only depends on $\sigma_t(x_t)$, where $\sigma_t(x_t) = \frac{1}{M}\sum_{i=1}^M x_t^i$ and we defined $x_t := [x_t^1, \dots, x_t^M]^\top \in \mathbb{R}^M$. The different components of x are in general still coupled by the individual constraint $\mathcal{X}^i$. The cost function (3.22) is widely used in the literature, especially for plug-in electric vehicles (see [MCH13, GPC16] and Section 6.1) and traffic networks (see [BMW56, CSSM04] and Section 6.2). While simple sufficient conditions for strong monotonicity and monotonicity for F_W can be directly derived by Lemma 2, the following Theorem 1 provides a more interesting sufficient condition for strong monotonicity of F_N.

Theorem 1. Let $\mathcal{X}$ be closed, convex. Assume that for each agent $i \in \{1, \dots, M\}$ the function v^i in (3.19) is convex and that for each t the price function p_t in (3.22) is twice continuously differentiable, strictly increasing. If there exists x^0 such that $\mathcal{X}^i \subseteq [0, x^0]^n$ for all $t \in \{1, \dots, n\}, i \in \{1, \dots, M\}$ and if

$$\min_{\substack{t\in\{1,\dots,n\}\\ z\in[0,x^0]}} \left(p_t'(z) - \frac{x^0 p_t''(z)}{8} \right) > 0. \tag{3.23}$$

then the operator F_N is strongly monotone.

Proof. Let us first show that p_t strictly increasing for all t implies p monotone:

$$(p(y) - p(z))^\top (y - z) = \sum_{t=1}^{n} (p_t(y_t) - p_t(z_t))(y_t - z_t) > 0.$$

Then by Lemma 2 the operator $F_\mathrm{W}(x) = [v^i(x^i) + p(\sigma(x))]_{i=1}^M$ is monotone. According to (3.20b), to show strong monotonicity of F_N it is sufficient to show that under condition (3.30) the term $[\nabla_z p(z)_{|z=\sigma(x)} x^i]_{i=1}^M$ is strongly monotone for all $x \in \mathcal{X}$, which is equivalent to $\nabla_x [\nabla_z p(z)_{|z=\sigma(x)} x^i]_{i=1}^M \succ 0$ for all $x \in \mathcal{X}$ by Proposition 8, due to compactness of $\mathcal{X}$. We have

$$\nabla_x [\nabla_z p(z)_{|z=\sigma(x)} x^i]_{i=1}^M = I_M \otimes \nabla_z p(z)_{|z=\sigma(x)} + \frac{1}{M} \mathbb{1}_M \otimes \left([\mathrm{diag}\{p_t''(\sigma_t) x_t^i\}_{t=1}^n]_{i=1}^M \right)^\top, \tag{3.24}$$

where $\mathrm{diag}\{p_t''(\sigma_t)x_t^i\}_{t=1}^n$ is the diagonal matrix whose entry in position (t,t) is $p_t''(\sigma_t)x_t^i$. The permutation matrix $P = [[\mathbf{e}_{t+(i-1)n}^\top]_{i=1}^M]_{t=1}^n$ (where $\mathbf{e}_i$ is the i^th vector of the canonical basis) permutes (3.24) in block-diagonal form

$$P \nabla_x [\nabla_z p(z)_{|z=\sigma(x)} x^i]_{i=1}^M P^\top = \begin{bmatrix} p_1'(\sigma_1) I_M & & \\ & \ddots & \\ & & p_n'(\sigma_n) I_M \end{bmatrix} + \frac{1}{M} \begin{bmatrix} p_1''(\sigma_1) x_1 \mathbb{1}_M^\top & & \\ & \ddots & \\ & & p_n''(\sigma_n) x_n \mathbb{1}_M^\top \end{bmatrix}, \tag{3.25}$$

where $x_t = [x_t^i]_{i=1}^M$. It suffices to show $p_t'(\sigma_t) I_M + \frac{1}{M} p_t''(\sigma_t) x_t \mathbb{1}_M^\top \succ 0$ for all t. By the following Lemma 3, if $x_t \in [0, x^0]$ then $\lambda_{\min}\left(x_t \mathbb{1}_M^\top + \mathbb{1}_M x_t^\top \right)/2 \geq -\frac{x^0 M}{8}$, which concludes the proof. $\square$

Lemma 3. For all $M \in \mathbb{N}$ it holds

$$\min_{y \in [0,1]^M} \lambda_{\min}\left(y \mathbb{1}_M^\top + \mathbb{1}_M y^\top \right) \geq -\frac{M}{4}. \tag{3.26}$$

Proof. The statement is trivially true for $M = 1$. For $M > 1$, the left-hand side of (3.26) is equivalent to

$$\min_{\substack{y\in[0,1]^M\\ \|v\|=1}} v^\top\big(y\mathbb{1}_M^\top + \mathbb{1}_M y^\top\big)\, v = \min_{\substack{y\in[0,1]^M\\ \|v\|=1}} 2\left(v^\top y\right)\left(\mathbb{1}_M^\top v\right). \tag{3.27}$$

Let us consider a pair $y^\star, v^\star$ minimizing (3.27). If $\mathbb{1}_M^\top v^\star = 0$, then the bound (3.26) is trivially satisfied. We are left with two cases, $\mathbb{1}_M^\top v^\star > 0$ and $\mathbb{1}_M^\top v^\star < 0$. Let us start analyzing $\mathbb{1}_M^\top v^\star > 0$. To minimize $2\left(v^\top y\right)\left(\mathbb{1}_M^\top v\right)$, it must be

$$y_i^\star = \begin{cases} 0 & \text{if } v_i^\star > 0 \\ 1 & \text{if } v_i^\star < 0, \end{cases} \quad \text{for all } i \in \{1,\dots,M\}. \tag{3.28}$$

Without loss of generality, we can assume $y_i^\star \in \{0,1\}$ if $v_i^\star = 0$. Hence $y^\star \in \{0,1\}^M$ and (3.26) reduces to

$$\min_{p\in\{0,\dots,M\}} \lambda_{\min}\left[\begin{array}{c|c} 2(\mathbb{1}_p\mathbb{1}_p^\top) & \mathbb{1}_p\mathbb{1}_{(M-p)}^\top \\ \hline \mathbb{1}_{(M-p)}\mathbb{1}_p^\top & \mathbb{0}_{(M-p)}\mathbb{0}_{(M-p)}^\top \end{array}\right], \tag{3.29}$$

where without loss of generality we assumed the first p components of $y^\star$ to be 1 and the remaining to be 0. Note that the matrix in (3.29) features p identical rows followed by $M-p$ other identical rows. Hence any of its eigenvectors must have p identical components followed by $M-p$ other identical components. With this observation and the definition of eigenvalue, simple algebraic computations show that the matrix in (3.29) has only two distinct eigenvalues, the minimum of the two being $p-\sqrt{Mp}$. The function $p-\sqrt{Mp}$ is minimized over the reals for $p = M/4$ with corresponding minimum $\lambda_{\min} = -M/4$, as it can be seen by using the change of variables $p = q^2$ and minimizing the quadratic function $q^2 - \sqrt{M}q$. Since $p \in \{0,\dots,M\}$ in (3.29), the value $-M/4$ is a lower bound for the minimum eigenvalue, and it is attained only if M is a multiple of 4. We conclude by noting that the derivation for the case $\mathbb{1}_M^\top v^\star < 0$ is identical to the derivation for the case $\mathbb{1}_M^\top v^\star > 0$ just shown, upon switching 0 and 1 in (3.28). □

Corollary 2. Under the assumptions of Theorem 1, if $p_t'' < 0$ for all t, then F_{N} is strongly monotone. □

Remark 1. The assumption in Theorem 1 of non-negativity of the agents' decisions x is met for several relevant applications, as the three studied in Chapter 6. If x can take negative values, in Theorem 1 we can assume $\mathcal{X}^i \subseteq [-x^0, x^0]^n$, rather than $\mathcal{X}^i \subseteq [0, x^0]^n$. Then to guarantee strong monotonicity of F_{N} condition (3.23) has to be replaced with

$$\min_{\substack{t\in\{1,\dots,n\}\\ z\in[-x^0,x^0]}} \left(p_t'(z) - x^0 p_t''(z)\right) > 0. \tag{3.30}$$

This follows from replacing 0 with -1 in (3.28), which results in the matrix (3.29) becoming $[2\mathbb{1}_p\mathbb{1}_p^\top, \mathbb{0}_{p\times M-p}; \mathbb{0}_{p\times M-p}, -2\mathbb{1}_{M-p}\mathbb{1}_{M-p}^\top]$, whose minimum eigenvalue is $-2M$, obtained for $p = 0$. □

To the best of our knowledge, the only work studying uniqueness of the variational Nash equilibrium for a diagonal price function is [YSM11], which also studies an aggregative game and in [YSM11, Lemma 3] it exploits expression (3.25) to give conditions for $\nabla_x F_{\mathrm{N}}(x)$ to be a P-matrix, which in turn guarantees uniqueness of the Nash equilibrium in absence of coupling constraints. It is interesting to note that uniqueness in [YSM11] holds under $p_t' > 0, p_t'' > 0$, whereas Theorem 1 guarantees uniqueness under the complementary case $p_t' > 0, p_t'' < 0$.

Affine price function

We now focus on a cost function which is a specific case of (3.19):

$$J^i(x^i, \sigma(x)) := \frac{1}{2}(x^i)^\top Q x^i + (C\sigma(x) + c^i)^\top x^i\,, \tag{3.31}$$

where $Q \in \mathbb{R}^{n\times n}$ is symmetric, $C \in \mathbb{R}^{n\times n}, c^i \in \mathbb{R}^n$. These cost functions have been used in many works on aggregative games [HCM07, GPC16, BP13]. Since the operators $F_{\mathrm{N}}, F_{\mathrm{W}}$ defined in (3.11) are obtained by differentiating quadratic functions, their expression is affine and can be explicitly characterized as

$$F_{\mathrm{W}}(x) = \left(I_M \otimes Q + \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top \otimes C\right) x + c, \tag{3.32a}$$

$$F_{\mathrm{N}}(x) = F_{\mathrm{W}}(x) + \frac{1}{M}(I_M \otimes C^\top)x, \tag{3.32b}$$

where $c = [c^1; \dots; c^M]$. The following lemma exploits the characterization (3.32) to derive sufficient conditions for strong monotonicity of F_{W}, F_{N}.

Lemma 4. The following hold.

- If $Q \succ 0$, $C \succeq 0$ or if $Q \succeq 0$, $C \succ 0$ then F_{N} in (3.32b) is strongly monotone.
- If $Q \succ 0$, $C \succeq 0$ then F_{W} in (3.32a) is strongly monotone.
- If $Q \succ 0$, $Q - C^\top Q^{-1} C \succ 0$ then F_{W} in (3.32a) is strongly monotone.

Proof. By Proposition 8, strong monotonicity of F_{W} in (3.32a) is equivalent to $\nabla_x F_{\mathrm{W}}(x) = \left(I_M \otimes Q + \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top \otimes C\right)^\top \succ 0$, which is independent from x. In the same way, strong monotonicity of F_{N} in (3.32b) is equivalent to $\left(I_M \otimes Q + \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top \otimes C\right)^\top + \frac{1}{M}(I_M \otimes C^\top)^\top \succ 0$. Building on this, the first two statements are straightforward to prove. The last statement needs a slightly more elaborate proof. By using Schur's theorem

$$\left.\begin{array}{l} Q \succ 0 \\ Q - C^\top Q^{-1} C \succ 0 \end{array}\right\} \Rightarrow \begin{bmatrix} Q & C^\top \\ C & Q \end{bmatrix} \succ 0 \Rightarrow \begin{bmatrix} x^\top & x^\top \end{bmatrix} \begin{bmatrix} Q & C^\top \\ C & Q \end{bmatrix} \begin{bmatrix} x \\ x \end{bmatrix} > 0,\ \forall\, x \in \mathbb{R}^n$$

$$\Leftrightarrow x^\top \left(Q + \frac{C + C^\top}{2}\right) x > 0,\ \forall\, x \in \mathbb{R}^n \Leftrightarrow Q + C \succ 0. \tag{3.33}$$

We show in the following that $Q+C \succ 0$ implies $\nabla_x F_{\mathrm{W}}(x) \succ 0$ by proving non-negativity of each eigenvalue λ_j of $\nabla_x F_{\mathrm{W}}(x) = I_M \otimes Q + \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top \otimes (C+C^\top)/2$. Corresponding to λ_j there exists an eigenvector $v_j \neq \mathbb{0}_{Mn}$ such that

$$\left(I_M \otimes Q + \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top \otimes \frac{C+C^\top}{2}\right) v_j = \lambda_j v_j \Leftrightarrow Q v_j^i + \frac{C+C^\top}{2M}\sum_{i=1}^M v_j^i = \lambda_j v_j^i,\ \forall\, i$$
$$\Rightarrow \left(Q + \frac{C+C^\top}{2}\right)\sum_{i=1}^M v_j^i = \lambda_j \sum_{i=1}^M v_j^i. \tag{3.34}$$

As a consequence, λ_j is an eigenvalue of $Q + (C+C^\top)/2$ (hence $\lambda_j > 0$ by (3.33)) if $\sum_{i=1}^M v_j^i \neq \mathbb{0}_n$. If instead $\sum_{i=1}^M v_j^i = \mathbb{0}_n$ then from (3.34) it follows that $Qv_j^i = \lambda_j v_j^i$, $\forall\, i$, so λ_j is an eigenvalue of Q, hence $\lambda_j > 0$ by assumption. □

The works [PKL16] and [BG17] study strong monotonicity of F_{N} for affine p. Specifically, [PKL16, Proposition 1] proves the first bullet, while slightly different necessary and sufficient conditions are provided in [BG17, Corollary 1].

3.4 Distance between variational Nash and Wardrop

It is clear from Lemma 2, Theorem 1 and Lemma 4 that often only one of F_{N} and F_{W} features strong monotonicity, which is required to guarantee that a variational equilibrium can be achieved using the algorithms proposed in Chapter 4. Hence it is important to derive results on the distance between the two variational equilibria, which is the goal of this section, where we focus on aggregative games with large number of agents. The basic intuition is that as the number of agents M grows, solving the VI for the variational Nash is similar to solving the VI for the variational Wardrop, because F_{N} converges to F_{W}, as it can be seen for instance in (3.20).

Formally, we consider a sequence of games $(\mathcal{G}_M)_{M=1}^\infty$. For fixed M, the game $\mathcal{G}_M$ is played among M agents and is defined as in (3.3) with an arbitrary coupling constraint $\mathcal{C}$ and, for every agent i, arbitrary $J^i(x^i, \sigma(x))$ and $\mathcal{X}^i$. For the sake of readability, we avoid the explicit dependence on M in denoting these quantities and in denoting x_{N}, x_{W}, F_{N}, F_{W}.

Assumption 2. There exists a convex, compact set $\mathcal{X}^0 \subset \mathbb{R}^n$ such that $\cup_{i=1}^M \mathcal{X}^i \subseteq \mathcal{X}^0$ for each $\mathcal{G}_M$ in the sequence $(\mathcal{G}_M)_{M=1}^\infty$. For each M and $i \in \{1, \dots, M\}$, the function $J^i(z_1, z_2)$ is Lipschitz with respect to z_2 in $\mathcal{X}^0$ with Lipschitz constant L_2 independent from M, i and $z_1 \in \mathcal{X}^i$. □

We note that Assumption 2 implies that $\sigma(x) \in \mathcal{X}^0$ for any M and any $x \in \mathcal{X}^1 \times \cdots \times \mathcal{X}^M$. Moreover, under Assumption 2 we define $R := \max_{y \in \mathcal{X}^0}\{\|y\|\}$. Furthermore, if the cost function (3.1) takes the specific form (3.19), then p being Lipschitz in $\mathcal{X}^0$ with constant L_p implies $J^i(z_1, z_2)$ being Lipschitz with respect to z_2 in $\mathcal{X}^0$ with constant $L_2 = RL_p$, because

$$\begin{aligned} \|J^i(z_1,z_2) - J^i(z_1,z_2')\| &= \|(p(z_2) - p(z_2'))^\top z_1\| \\ &\overset{\substack{\text{Cauchy}\\\text{Schwarz}}}{\leq} \|p(z_2) - p(z_2')\|\|z_1\| \leq RL_p\|z_2 - z_2'\|. \end{aligned} \tag{3.35}$$

The next proposition shows that every Wardrop equilibrium is an ε-Nash equilibrium, with ε tending to zero as M grows.

Proposition 16 ([GPC16, Theorem 1])**.** Let the sequence of games $(\mathcal{G}_M)_{M=1}^\infty$ satisfy Assumption 2. For each $\mathcal{G}_M$, every Wardrop equilibrium is an ε-Nash equilibrium, with $\varepsilon = \frac{2RL_2}{M}$. □

Proof. Consider any Wardrop equilibrium x_W of $\mathcal{G}_M$ (not necessarily a variational one). By Definition 10, $x_\mathrm{W} \in \mathcal{Q}$ and for each agent i

$$J^i(x^i_\mathrm{W}, \sigma(x_\mathrm{W})) \leq J^i(x^i, \sigma(x_\mathrm{W})), \qquad \forall\, x^i \in \mathcal{Q}^i(x^{-i}_\mathrm{W}).$$

It follows that for each agent i and for all $x^i \in \mathcal{Q}^i(x^{-i}_\mathrm{W})$

$$\begin{aligned} &J^i(x^i_\mathrm{W}, \sigma(x_\mathrm{W})) - J^i(x^i, \tfrac{1}{M}(x^i + \textstyle\sum_{j\neq i} x^j_\mathrm{W})) \\ &= \underbrace{J^i(x^i_\mathrm{W}, \sigma(x_\mathrm{W})) - J^i(x^i, \sigma(x_\mathrm{W}))}_{\leq 0} + J^i(x^i, \sigma(x_\mathrm{W})) - J^i(x^i, \tfrac{1}{M}(x^i + \textstyle\sum_{j\neq i} x^j_\mathrm{W})) \\ &\leq L_2\|\sigma(x_\mathrm{W}) - \tfrac{1}{M}(x^i + \textstyle\sum_{j\neq i} x^j_\mathrm{W})\| = \tfrac{L_2}{M}\|(x^i_\mathrm{W} + \textstyle\sum_{j\neq i} x^j_\mathrm{W}) - (x^i + \textstyle\sum_{j\neq i} x^j_\mathrm{W})\| \\ &= \tfrac{L_2}{M}\|x^i_\mathrm{W} - x^i\| \leq \tfrac{2RL_2}{M}, \end{aligned}$$

hence x_W is an ε-Nash equilibrium of $\mathcal{G}_M$, by Definition 9. □

Proposition 16 is a strong result but it provides no information on the distance between the set of strategies constituting a Nash and the set of strategies constituting a Wardrop equilibrium. In the following we study this distance for variational equilibria.

Theorem 2. Let the sequence of games $(\mathcal{G}_M)_{M=1}^\infty$ satisfy Assumption 2, and each $\mathcal{G}_M$ satisfy Assumption 1. Then the following hold.

1. If the operator F_N relative to $\mathcal{G}_M$ is strongly monotone on $\mathcal{Q}$ with monotonicity constant $\alpha_M > 0$, then there exists a unique variational Nash equilibrium $\bar{x}_\mathrm{N}$ of $\mathcal{G}_M$. Moreover, for any variational Wardrop equilibrium $\bar{x}_\mathrm{W}$

$$\|\bar{x}_\mathrm{N} - \bar{x}_\mathrm{W}\| \leq \frac{L_2}{\alpha_M \sqrt{M}}. \tag{3.36}$$

   As a consequence, if $\alpha_M\sqrt{M} \to \infty$ as $M \to \infty$, then $\|\bar{x}_\mathrm{N} - \bar{x}_\mathrm{W}\| \to 0$ as $M \to \infty$.

2. If the operator F_{W} relative to $\mathcal{G}_M$ is strongly monotone on $\mathcal{Q}$ with monotonicity constant $\alpha_{_M} > 0$, then there exists a unique variational Wardrop equilibrium $\bar{x}_{\mathrm{W}}$ of $\mathcal{G}_M$. Moreover, for any variational Nash equilibrium $\bar{x}_{\mathrm{N}}$

$$\|\bar{x}_{\mathrm{N}} - \bar{x}_{\mathrm{W}}\| \leq \frac{L_2}{\alpha_{_M}\sqrt{M}}. \tag{3.37}$$

   As a consequence, if $\alpha_{_M}\sqrt{M} \to \infty$ as $M \to \infty$, then $\|\bar{x}_{\mathrm{N}} - \bar{x}_{\mathrm{W}}\| \to 0$ as $M \to \infty$.

3. If in each game $\mathcal{G}_M$ the cost function $J^i(x^i, \sigma(x))$ takes the form (3.19) with $v^i = 0$ and p is strongly monotone in $\mathcal{X}^0$ with monotonicity constant α, then there exists a unique $\bar{\sigma}$ such that $\sigma(\bar{x}_{\mathrm{W}}) = \bar{\sigma}$ for any variational Wardrop equilibrium $\bar{x}_{\mathrm{W}}$ of $\mathcal{G}_M$. Moreover, for any variational Nash equilibrium $\bar{x}_{\mathrm{N}}$ of $\mathcal{G}_M$ and for any variational Wardrop equilibrium[1] $\bar{x}_{\mathrm{W}}$ of $\mathcal{G}_M$

$$\|\sigma(\bar{x}_{\mathrm{N}}) - \sigma(\bar{x}_{\mathrm{W}})\| \leq \sqrt{\frac{2RL_2}{\alpha M}}. \tag{3.38}$$

   Hence, $\|\sigma(\bar{x}_{\mathrm{N}}) - \sigma(\bar{x}_{\mathrm{W}})\| \to 0$ as $M \to \infty$. □

Proof. 1) We first bound the distance between the operators F_{N} and F_{W} in terms of M. By (3.11) it holds

$$\begin{aligned}\|F_{\mathrm{N}}(x) - F_{\mathrm{W}}(x)\|^2 &= \|[\nabla_{x^i} J^i(x^i, \sigma(x))]_{i=1}^M - [\nabla_{x^i} J^i(x^i, z)_{|z=\sigma(x)}]_{i=1}^M\|^2 \\ &= \sum_{i=1}^M \|\frac{1}{M}\nabla_z J^i(x^i, z)_{|z=\sigma(x)}\|^2 \leq \frac{1}{M^2}\sum_{i=1}^M L_2^2 = \frac{L_2^2}{M},\end{aligned}$$

where the inequality follows from the fact that $J^i(z_1, z_2)$ is Lipschitz in z_2 in $\mathcal{X}^0$ with constant L_2 by Assumption 2 and hence the term $\|\nabla_z J^i(x^i, z)_{|z=\sigma(x)}\|$ is bounded by L_2 by definition of derivative. It follows that

$$\|F_{\mathrm{N}}(x) - F_{\mathrm{W}}(x)\| \leq \frac{L_2}{\sqrt{M}}. \tag{3.39}$$

for all $x \in \mathcal{X}^0$. We exploit (3.39) to bound the distance between Nash and Wardrop strategies. Since F_{N} is strongly monotone on $\mathcal{Q}$ by assumption, $\mathrm{VI}(\mathcal{Q}, F_{\mathrm{N}})$ has a unique solution $\bar{x}_{\mathrm{N}}$ by Proposition 7. Moreover, by [Nag13, Theorem 1.14] for all solutions $\bar{x}_{\mathrm{W}}$ of $\mathrm{VI}(\mathcal{Q}, F_{\mathrm{W}})$ it holds

$$\|\bar{x}_{\mathrm{N}} - \bar{x}_{\mathrm{W}}\| \leq \frac{1}{\alpha_{_M}}\|F_{\mathrm{N}}(\bar{x}_{\mathrm{W}}) - F_{\mathrm{W}}(\bar{x}_{\mathrm{W}})\|. \tag{3.40}$$

The bound (3.40) can be thought of as the generalization to VI of the same bound for strongly convex functions. Combining this with equation (3.39) yields the result.

[1] If p is Lipschitz with constant L_p, then in (3.38) L_2 can be replaced by RL_p, as by (3.35). This is used in the application Sections 6.1, 6.2.

2) As in the above, with Nash in place of Wardrop and viceversa.
3) Any solution $\bar{x}_\mathrm{W}$ to VI(Q, F_W) satisfies

$$\begin{aligned} &F_\mathrm{W}(\bar{x}_\mathrm{W})^\top (x - \bar{x}_\mathrm{W}) \geq 0,\ \forall x \in Q \Leftrightarrow \textstyle\sum_{i=1}^M p(\sigma(\bar{x}_\mathrm{W}))^\top (x^i - \bar{x}^i_\mathrm{W}) \geq 0,\ \forall x \in Q \Leftrightarrow \\ &p(\sigma(\bar{x}_\mathrm{W}))^\top (\sigma(x) - \sigma(\bar{x}_\mathrm{W})) \geq 0,\ \forall x \in Q. \end{aligned} \tag{3.41}$$

Any solution $\bar{x}_\mathrm{N}$ to VI(Q, F_N) satisfies

$$\begin{aligned} &F_\mathrm{N}(\bar{x}_\mathrm{N})^\top (x - \bar{x}_\mathrm{N}) \geq 0,\ \forall x \in Q \Leftrightarrow \\ &p(\sigma(\bar{x}_\mathrm{N}))^\top (\sigma(x) - \sigma(\bar{x}_\mathrm{N})) + \frac{1}{M^2} \sum_{i=1}^M (\nabla_z p(z)_{|z=\sigma(\bar{x}_\mathrm{N})} \bar{x}^i_\mathrm{N})^\top (x^i - \bar{x}^i_\mathrm{N}) \geq 0,\ \forall x \in Q. \end{aligned} \tag{3.42}$$

Exploiting the strong monotonicity of p in $\mathcal{X}^0$, one has

$$\begin{aligned} &\alpha \|\sigma(\bar{x}_\mathrm{W}) - \sigma(\bar{x}_\mathrm{N})\|^2 \leq (p(\sigma(\bar{x}_\mathrm{W})) - p(\sigma(\bar{x}_\mathrm{N})))^\top (\sigma(\bar{x}_\mathrm{W}) - \sigma(\bar{x}_\mathrm{N})) \\ &= p(\sigma(\bar{x}_\mathrm{W}))^\top (\sigma(\bar{x}_\mathrm{W}) - \sigma(\bar{x}_\mathrm{N})) - p(\sigma(\bar{x}_\mathrm{N}))^\top (\sigma(\bar{x}_\mathrm{W}) - \sigma(\bar{x}_\mathrm{N})) \\ &\underset{\text{by (3.41)}}{\leq} -p(\sigma(\bar{x}_\mathrm{N}))^\top (\sigma(\bar{x}_\mathrm{W}) - \sigma(\bar{x}_\mathrm{N})) \underset{\text{by (3.42)}}{\leq} \frac{1}{M^2} \sum_{i=1}^M (\bar{x}^i_\mathrm{N})^\top (\nabla_z p(z)_{|z=\sigma(\bar{x}_\mathrm{N})})^\top (\bar{x}^i_\mathrm{W} - \bar{x}^i_\mathrm{N}) \\ &= \tfrac{1}{M^2} \textstyle\sum_{i=1}^M (\bar{x}^i_\mathrm{N})^\top (\nabla_z J^i(\bar{x}^i_\mathrm{W}, z)_{|z=\sigma(\bar{x}_\mathrm{N})} - \nabla_z J^i(\bar{x}^i_\mathrm{N}, z)_{|z=\sigma(\bar{x}_\mathrm{N})}) \\ &\leq \tfrac{1}{M^2} \textstyle\sum_{i=1}^M \|\bar{x}^i_\mathrm{N}\| \|\nabla_z J^i(\bar{x}^i_\mathrm{W}, z)_{|z=\sigma(\bar{x}_\mathrm{N})}\| + \tfrac{1}{M^2} \sum_{i=1}^M \|\bar{x}^i_\mathrm{N}\| \|\nabla_z J^i(\bar{x}^i_\mathrm{N}, z)_{|z=\sigma(\bar{x}_\mathrm{N})}\| \\ &\leq \tfrac{2L_2}{M^2} \textstyle\sum_{i=1}^M \|\bar{x}^i_\mathrm{N}\| \leq \tfrac{2L_2}{M^2} \sum_{i=1}^M R \leq \tfrac{1}{M} 2RL_2. \end{aligned}$$

We conclude that $\|\sigma(\bar{x}_\mathrm{W}) - \sigma(\bar{x}_\mathrm{N})\| \leq \sqrt{\frac{2RL_2}{\alpha M}}$. □

The bound (3.36) implies that $\|\bar{x}_\mathrm{W} - \bar{x}'_\mathrm{W}\| \leq \frac{2L_p R}{\alpha_M \sqrt{M}}$ for any two variational Wardrop equilibria $\bar{x}_\mathrm{W}$ and $\bar{x}'_\mathrm{W}$ of $\mathcal{G}_M$. Hence one can draw conclusions on the distance between variational Wardrop equilibria without assumptions on the corresponding operator F_W, but only on F_N. For instance, if $\alpha_M \sqrt{M} \to \infty$ as $M \to \infty$, then any two variational Wardrop equilibria tend to each other. In the same way, (3.37) implies that for any two variational Nash equilibria $\bar{x}_\mathrm{N}$ and $\bar{x}'_\mathrm{N}$ of $\mathcal{G}_M$ it holds $\|\bar{x}_\mathrm{N} - \bar{x}'_\mathrm{N}\| \leq \frac{2L_p R}{\alpha_M \sqrt{M}}$.

The first two statements of Theorem 2 assume strong monotonicity of either F_W, F_N. Due to the term $\frac{1}{M}[\nabla_z p(z) x^i{}_{|z=\sigma(x)}]_{i=1}^M$ in (3.20b), in general assessing monotonicity of F_N is more difficult than assessing monotonicity of F_W. The next lemma gives a sufficient condition for strong monotonicity of F_N based on the strong monotonicity of F_W.

Lemma 5. Let the sequence of games $(\mathcal{G}_M)_{M=1}^\infty$ satisfy Assumption 2 and for each game $\mathcal{G}_M$ let the cost functions be of the form (3.19). Moreover, assume that for each $\mathcal{G}_M$ the corresponding F_W is strongly monotone with monotonicity constant α_W independent from M and that p is $\mathcal{C}^2$ in $\mathcal{X}^0$. Then for any $\alpha_\mathrm{N} < \alpha_\mathrm{W}$ there exists $\hat{M}$ such that F_N is strongly monotone with monotonicity constant α_N, for each $M > \hat{M}$.

Proof. Let us use within this proof the notation $D(x) := \frac{1}{M}[\nabla_z p(z) x^i{}_{|z=\sigma(x)}]_{i=1}^M$, hence $D(x)$ is the difference between $F_\mathrm{N}(x)$ and $F_\mathrm{W}(x)$, as by (3.20). We start by bounding the Lipschitz constant of D in terms of M. Since p is $\mathcal{C}^2$ on the compact set $\mathcal{X}^0$, there exists $L_{pp} > 0$ such that $\sum_{j=1}^n \|[\nabla_z([\nabla_z p(z)]_{(:,j)} x^j)]_{|z=\sigma(x)}\|_2 \leq L_{pp}, \forall x \in \mathcal{X}^0$. Moreover,

$$[M\nabla_x D(x)]^\top = I_M \otimes \nabla_z p(z)_{|z=\sigma(x)} + \frac{1}{M}\begin{bmatrix} \nabla_z(\nabla_z p(z) x^1) & & \nabla_z(\nabla_z p(z) x^1) \\ & \ddots & \\ \nabla_z(\nabla_z p(z) x^M) & & \nabla_z(\nabla_z p(z) x^M) \end{bmatrix}_{|z=\sigma(x)}.$$

Hence

$$\begin{aligned}
&\|[M\nabla_x D(x)]^\top\|_2 \leq \sqrt{Mn}\|[M\nabla_x D(x)]^\top\|_\infty \\
&\leq \sqrt{Mn}\big(\|\nabla_z p(z)_{|z=\sigma(x)}\|_\infty + \max_i \|[\nabla_z(\nabla_z p(z) x^i)]_{|z=\sigma(x)}\|_\infty\big) \\
&\leq \sqrt{Mn}\big(\|\nabla_z p(z)_{|z=\sigma(x)}\|_2 + \max_i \|[\nabla_z(\nabla_z p(z) x^i)]_{|z=\sigma(x)}\|_2\big) \\
&\leq \sqrt{M}n\,(L_p + L_{pp}R)\,.
\end{aligned}$$

This implies $\|[\nabla_x D(x)]^\top\|_2 \leq \frac{n(L_p + L_{pp}R)}{\sqrt{M}} =: \frac{L_d}{\sqrt{M}}$, meaning that D is Lipschitz with constant $\frac{L_d}{\sqrt{M}}$. Consequently,

$$\begin{aligned}
&(F_\mathrm{N}(x) - F_\mathrm{N}(y))^\top(x-y) = (F_\mathrm{W}(x) - F_\mathrm{W}(y))^\top(x-y) + (D(x) - D(y))^\top(x-y) \\
&\geq \alpha_\mathrm{W}\|x-y\|^2 - \|D(x) - D(y)\|\|x-y\| \geq (\alpha_\mathrm{W} - \tfrac{L_d}{\sqrt{M}})\|x-y\|^2.
\end{aligned}$$

The statement is proven upon noticing that for any $\alpha_\mathrm{N} < \alpha_\mathrm{W}$ there exists $\hat{M}$ such that $(\alpha_\mathrm{W} - \frac{L_d}{\sqrt{M}}) > \alpha_\mathrm{N} > 0$. □

In Lemma 5 the role of Nash and Wardrop is symmetric and it also holds that strong monotonicity of F_N implies strong monotonicity of F_W for M large enough.

Comparison with the literature

Proposition 16 states that, under fairly general assumptions, any Wardrop equilibrium is an ε-Nash equilibrium. Such result follows directly from the fact that each agent contributes only via the average and that the cost functions are Lipschitz. Consequently, the contribution of each agent scales linearly with the inverse of the number of agents M. This same idea is used to prove similar results in many previous contributions. Regarding the Wardrop equilibrium expressed in terms of $\sigma(x)$ as in (3.10), the case of potential games is investigated in [AW04, ABEA06], routing games are considered in [ACA11], flow control and routing in communication networks are discussed in [ABS02b]. The statement of Proposition 16 is also showed with a more involved proof in [GPC16]. Proposition 16 is a trivial extension of those works to generic aggregative games with coupling constraints.

Our main result is to prove that, by introducing further assumptions, one can actually go beyond Proposition 16 and derive bounds on the Euclidean distance between Nash and Wardrop equilibria. In Theorem 2 we consider two types of additional assumptions: the first is strong monotonicity of either the Nash or Wardrop operator (statements 1 and 2), the second is a structural assumption on the cost functions (statement 3). The only previous results bounding the Euclidean distance between the two equilibria that we are aware of are obtained in [HM85]. Therein a similar bound to our result of Theorem 2-3) is derived specific to routing/congestion games. However, that work assumes that the number of agents increases by means of identical replicas of the agents. To be specific, [HM85] considers an original game with $\tilde{M}$ agents with constraints $\{\mathcal{X}^i\}_{i=1}^{\tilde{M}}$ and cost functions $\{J^i\}_{i=1}^{\tilde{M}}$; each game $\mathcal{G}_M$ features $M \cdot \tilde{M}$ agents, where each original agent (i.e., its constraints and cost functions) is replicated M times. We here prove that a similar argument as in [HM85] can be used to address the case of generic new agents instead of identical copies. Moreover, the results in Theorem 2-1) and Theorem 2-2) address a more general class of aggregative games (i.e. not necessarily congestion games) by employing a new type of argument, based on a sensitivity analysis result for variational inequalities with perturbed strongly monotone operators [Nag13, Theorem 1.14]. We note that the works [DN87, AW04, ABEA06] guarantee convergence of Nash to Wardrop in terms of Euclidean distance, but do not provide a bound on the convergence rate.

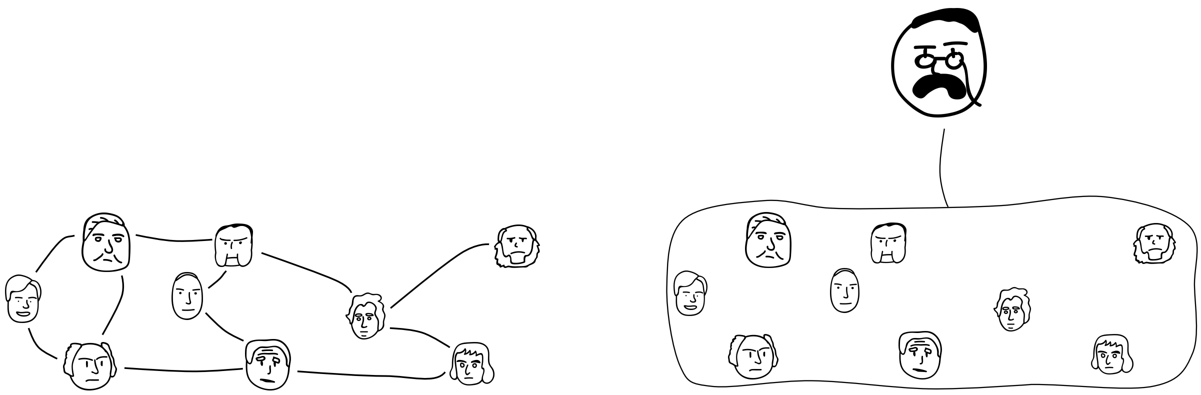
Finally, our results are derived for variational equilibria. We remark that if there are no coupling constraints, as in the previous works, then any equilibrium is a variational equilibrium. Hence our results subsume the results above. We remark that including coupling constraints does not increase the complexity of the mathematical treatment of Section 3.4; on the contrary the design of the algorithms in Chapters 4 and 5 is specifically tailored to account for coupling constraints.

CHAPTER 4

Two parallel algorithms

While the main result of Chapter 3 is on the relations between Nash and Wardrop equilibria, Chapters 4 and 5 propose algorithms that converge to such equilibria. We make a distinction between distributed and parallel information structure:

- in distributed algorithms each agent can only exchange information with his neighbors, as specified by an underlying communication network (see Figure 4.1a);
- in parallel algorithms the agents do not communicate with each other, but there exists a central operator (or facilitator) which can measure aggregate quantities relative to the agents and broadcast information to the them. For instance, if the cost function is as in (3.19), the central operator can measure the average strategy $\sigma(x)$ and broadcast the price $p(\sigma(x))$.



(a) Distributed (b) Parallel

Figure 4.1: Distributed (a) and parallel (b) information exchange.

In large-scale applications parallel or distributed schemes are often preferable to centralized ones for reasons of privacy and computational tractability. A thorough discussion of advantages and disadvantages of parallel and distributed computation is outside the scope of this thesis and can be found in [BT97, Section 1.1]. Chapter 4 proposes two parallel algorithms and Chapter 5 proposes a distributed algorithm.

In the previous Chapter 3 we focused on a sequence of games with increasing number of agents, because we studied proximity between variational Nash and Wardrop equilibria for large number of agents. Here instead we propose algorithms for finding the variational equilibria of the game $\mathcal{G}$ in (3.3) with fixed number of agents M. We start by making an assumption on the individual and coupling constraint sets.

Assumption 3. The coupling constraint in (3.2) is of the form

$$x \in \mathcal{C} := \{x \in \mathbb{R}^{Mn} \,|\, Ax \leq b\} \subset \mathbb{R}^{Mn}, \tag{4.1}$$

with $A := [A_1, \dots, A_M] \in \mathbb{R}^{m \times Mn}$, $A_i \in \mathbb{R}^{m \times n}$ for all $i \in \{1, \dots, M\}$, $b \in \mathbb{R}^m$. Moreover, for all $i \in \{1, \dots, M\}$, the set $\mathcal{X}^i$ in (3.3) can be expressed as $\mathcal{X}^i = \{x^i \in \mathbb{R}^n | g^i(x^i) \leq \mathbb{0}_{m_i}, h^i(x^i) = \mathbb{0}_{p_i}\}$, where $g^i : \mathbb{R}^n \to \mathbb{R}^{m_i}$ is continuously differentiable and $h^i : \mathbb{R}^n \to \mathbb{R}^{p_i}$ is affine. The set Q in (3.4), which can thus be expressed as $\mathcal{Q} = \{x \in \mathbb{R}^{Mn} | g^i(x^i) \leq \mathbb{0}_{m_i}, h^i(x^i) = \mathbb{0}_{p_i} \, \forall i, \, Ax \leq b\}$, satisfies Slater's constraint qualification of Definition 2. ■

Linearity of the coupling constraints arises in a range of applications, as explained in [FK07, page 188] and in [YP17, Remark 3.1]. Moreover, in the three applications of Chapter 6 the coupling constraints are linear. We also assume that agent i does not wish to disclose information about his cost function J^i and individual constraint set $\mathcal{X}^i$ and that he knows his influence on the coupling constraint, that is, the sub-matrix A_i in (4.1). Moreover, we assume the presence of a central operator that is able to measure the agents' average strategy $\sigma(x)$, to evaluate the quantity $Ax-b$ in (4.1) and to broadcast aggregate information to the agents. Based on this information structure, in the following we focus on the design of parallel algorithms to obtain a solution of either $\text{VI}(\mathcal{Q}, F_\text{N})$ or $\text{VI}(\mathcal{Q}, F_\text{W})$. As the techniques are the same for Nash and Wardrop equilibrium, we consider the general problem $\text{VI}(\mathcal{Q}, F)$, where F can be replaced with F_N or F_W. The symbol $F^i(x)$ indicates the i^th block of $F(x)$ which for $F = F_\text{N}$ equals $\nabla_{x^i} J^i(x^i, \sigma(x))$ and for $F = F_\text{W}$ equals $\nabla_{x^i} J^i(x^i, z)_{|z=\sigma(x)}$, so that $F(x) = [F^1(x); \dots; F^M(x)]$.

We start by noting that, if F is a monotone gradient operator on $\mathcal{Q}$, that is, if there exists a convex function $f(x) : \mathbb{R}^{Mn} \to \mathbb{R}$ such that $F(x) = \nabla_x f(x)$ for all $x \in \mathcal{Q}$, then by Figure 2.1 (and the explanation below it) $\text{VI}(\mathcal{Q}, F)$ is equivalent to the convex optimization program

$$\operatorname*{argmin}_{x \in \mathcal{Q}} \; f(x). \tag{4.2}$$

Therefore a solution of $\text{VI}(\mathcal{Q}, F)$ and thus a variational equilibrium can be found by applying any of the parallel optimization algorithms available in the literature [BT97] to problem (4.2); the parallel structure arises because each agent can evaluate $\nabla_{x^i} f(x)$ by knowing only his strategy x^i and $\sigma(x)$. Equivalently, if F is a gradient operator then $\mathcal{G}$ is a *potential game* [MS96] with potential function $f(x)$, hence parallel convergence tools available for potential games can also be employed [DHZ06, MAS09]. We anticipate that in the three applications of Chapter 6 the Wardrop operator F_W in (3.20a) is a gradient operator but the Nash operator F_N in (3.20b) is not.

In the following we intend to find a solution of $\mathrm{VI}(\mathcal{Q}, F)$ when F is not necessarily a gradient operator, so that these standard methods cannot be applied. Based on [SPF12], in order to propose parallel schemes in presence of coupling constraints we introduce two reformulations of $\mathrm{VI}(\mathcal{Q}, F)$ in an extended space $[x; \lambda]$ where λ is the dual variable relative to the coupling constraint $\mathcal{C}$. These two reformulations will then be used to propose two alternative algorithms. Specifically, we define for any $\lambda \in \mathbb{R}^m_{\geq 0}$ the game

$$\mathcal{G}(\lambda) := \begin{cases} \text{agents :} & \{1, \dots, M\} \\ \text{cost of agent } i : & J^i(x^i, \sigma(x)) + \lambda^\top A_i x^i \\ \text{individual constraint :} & \mathcal{X}^i \\ \text{coupling constraint :} & \mathbb{R}^{Mn}. \end{cases}$$

The fact that the coupling constraint $\mathcal{C}$ is equal to $\mathbb{R}^{Mn}$ means that $\mathcal{G}(\lambda)$ is a game without coupling constraints. We also introduce the extended $\mathrm{VI}(\mathcal{Y}, T)$ with $\mathcal{Y} \subseteq \mathbb{R}^{Mn+m}$ and $T : \mathbb{R}^{Mn+m} \to \mathbb{R}^{Mn+m}$ defined as

$$\mathcal{Y} := \mathcal{X} \times \mathbb{R}^m_{\geq 0}, \quad T(x, \lambda) := \begin{bmatrix} F(x) + A^\top \lambda \\ -(Ax - b) \end{bmatrix}. \tag{4.3}$$

The following proposition draws a connection between $\mathrm{VI}(\mathcal{Q}, F)$, the game $\mathcal{G}(\lambda)$ and $\mathrm{VI}(\mathcal{Y}, T)$.

Proposition 17. [SPF12, Section 4.3.2] Let Assumptions 1 and 3 hold. The following statements are equivalent.

1. The vector $\bar{x}$ is a solution of $\mathrm{VI}(\mathcal{Q}, F)$.
2. There exists $\bar{\lambda} \in \mathbb{R}^m_{\geq 0}$ such that $\bar{x}$ is a variational equilibrium of $\mathcal{G}(\bar{\lambda})$ and $\mathbb{0}_m \leq \bar{\lambda} \perp b - A\bar{x} \geq \mathbb{0}_m$.
3. There exists $\bar{\lambda} \in \mathbb{R}^m_{\geq 0}$ such that the vector $[\bar{x}; \bar{\lambda}]$ is a solution of $\mathrm{VI}(\mathcal{Y}, T)$. □

Proof. The proof consists in writing the KKT system relative to each of the statements and showing that the three KKT systems coincide.

1) Under Assumption 3 the set $\mathcal{Q}$, and consequently the sets $\{\mathcal{X}^i\}_{i=1}^M$, $\mathcal{X}$ and $\mathcal{Y}$, satisfy Slater's constraint qualification. As a consequence, by Proposition 4 $\mathrm{VI}(\mathcal{Q}, F)$ is equivalent to its KKT system

$$\begin{aligned} & F^i(x) + \nabla_{x^i} g^i(x^i)\mu^i + \nabla_{x^i} h^i(x^i)\nu^i + A_i^\top \lambda = \mathbb{0}_n, && \forall i \in \{1, \dots, M\} \quad (4.4a) \\ & \mathbb{0}_{m_i} \leq \mu^i \perp g^i(x) \leq \mathbb{0}_{m_i}, && \forall i \in \{1, \dots, M\} \quad (4.4b) \\ & h^i(x^i) = \mathbb{0}_{p_i}, && \forall i \in \{1, \dots, M\} \quad (4.4c) \\ & \mathbb{0}_m \leq \lambda \perp Ax - b \leq \mathbb{0}_m, && (4.4d) \end{aligned}$$

where μ^i is the dual variable associated to the constraint $g^i(x^i) \leq \mathbb{0}_{m_i}$, ν^i is the dual variable associated to the constraint $h(x) = \mathbb{0}_{p_i}$ and λ is the dual variable associated to the constraint $Ax \leq b$.

2) By definition of $\bar{x}$ being a variational equilibrium of $\mathcal{G}(\bar{\lambda})$, $\bar{x}$ solves VI$(\mathcal{Q}, F(x) + A^\top \bar{\lambda})$, which is equivalent to its KKT system:

$$F^i(x) + A_i^\top \bar{\lambda} + \nabla_{x^i} g^i(x^i)\mu^i + \nabla_{x^i} h^i(x^i)\nu^i = \mathbb{0}_n, \qquad \forall\, i \in \{1,\dots,M\} \quad (4.5a)$$
$$\mathbb{0}_{m_i} \leq \mu^i \perp g^i(x) \leq \mathbb{0}_{m_i}, \qquad \forall\, i \in \{1,\dots,M\} \quad (4.5b)$$
$$h^i(x^i) = \mathbb{0}_{p_i}, \qquad \forall\, i \in \{1,\dots,M\}. \quad (4.5c)$$

Moreover, by the second part of statement 2), $\bar{x}$ satisfies

$$\mathbb{0}_m \leq \bar{\lambda} \perp Ax - b \leq \mathbb{0}_m. \quad (4.5d)$$

3) Finally, by Proposition 4, VI$(\mathcal{Y}, T)$ is equivalent to its KKT system

$$F^i(x) + \nabla_{x^i} g^i(x^i)\mu^i + \nabla_{x^i} h^i(x^i)\nu^i + A_i^\top \lambda = \mathbb{0}_n, \qquad \forall\, i \in \{1,\dots,M\} \quad (4.6a)$$
$$-Ax + b - \lambda = \mathbb{0}_m, \quad (4.6b)$$
$$\mathbb{0}_{m_i} \leq \mu^i \perp g^i(x) \leq \mathbb{0}_{m_i}, \qquad \forall\, i \in \{1,\dots,M\} \quad (4.6c)$$
$$h^i(x^i) = \mathbb{0}_{p_i}, \qquad \forall\, i \in \{1,\dots,M\} \quad (4.6d)$$
$$\mathbb{0}_m \leq \eta \perp -\lambda \leq \mathbb{0}_m. \quad (4.6e)$$

where the primal variables are x and λ, while the dual variables are μ, ν and η. Substituting (4.6b) into (4.6e) results in $\mathbb{0}_m \leq \bar{\lambda} \perp Ax - b \leq \mathbb{0}_m$; one can then see by direct inspection that the KKT systems (4.4), (4.5) and (4.6) are equivalent. □

In subsection 4.1 we exploit the equivalence between 1) and 2) to propose a best-response algorithm that converges to a Wardrop equilibrium. In subsection 4.2 we leverage on the equivalence between 1) and 3) to propose a gradient step algorithm that can be employed to converge to a Nash equilibrium (with F_{N}) or to converge to a Wardrop equilibrium (with F_{W}).

	Nash	Wardrop
best response	-	Algorithm 3
gradient step	Algorithm 4	Algorithm 4

The core difference between a best-response algorithm and a gradient step one is that in the former each agent updates his strategy by minimizing his cost function, while in the latter each agent updates his strategy by performing one step in the steepest descent direction of his cost function. Whether it is more realistic to assume that one agent performs a complete minimization or moves precisely according to his gradient is in general application dependent and a proper discussion of the topic is beyond the scope of this thesis. We limit ourselves to pointing out that most works proposing algorithms to achieve Nash or Wardrop equilibria employ either best-response [SPF10, GPC16, Jen10, ABS02a, MCH13, Gra17] or gradient step [CLL14, Ros65, YSM11, YP17, KNS12, ZF16] schemes. The work [FK07] provides an overview of different paradigms, starting from best-response and gradient methods but covering also less common schemes. We perform a specific comparison with existing algorithms at the end of Section 4.1 and Section 4.2.

4.1 Best-response algorithm for Wardrop equilibrium

Based on the equivalence between 1) and 2) in Proposition 17, we here introduce Algorithm 3 to achieve a Wardrop equilibrium. The algorithm features an outer loop, in which the central operator broadcasts the dual variable $\lambda_{(k)}$ based on the current constraint violation, and an inner loop, in which the agents update their strategies to the Wardrop equilibrium of the game $\mathcal{G}(\lambda_{(k)})$ by using a best-response iteration. With the goal of enforcing the coupling constraint, in the inner loop the central operator provides an incentive / penalty, which is represented by the additive term $A_i^\top \lambda_{(k)} x^i$ in the cost of each agent.

Regarding the information exchange, the best-response iteration of the inner loop is parallel, as at every inner iteration the central operator broadcasts to all the agents the quantity $z_{(h)}$, which he can compute by measuring the average agents' strategy; in turn each agent optimizes his strategy given $z_{(h)}$. If the cost function is of the form (3.19), then the central operator can broadcast directly the price $p(z_{(h)})$ rather than $z_{(h)}$. In the outer loop, the central operator measures the coupling constraint violation to update the dual variable $\lambda_{(k)}$, which he then broadcasts to all the agents. If the coupling constraint is expressed on the average agents' strategy $\sigma(x)$, then the central operator only needs to measure $\sigma(x)$ to perform the dual update and in this case the entire algorithm is parallel.

We use the term *best-response* with a slight abuse of terminology, as in its usual meaning [SPF12, Section 4.2.4] this refers to $\mathrm{argmin}_{x^i \in \mathcal{X}^i} J^i(x^i, \frac{1}{M}(x^i + \sum_{j \neq i} x^j_{(h)})) + \lambda_{(k)}^\top A_i x^i$; here the term $\frac{1}{M}(x^i + \sum_{j \neq i} x^j_{(h)})$ is replaced with $z_{(h)}$, meaning that at each iteration the agent considers the average strategy as fixed and provided by the central operator, who updates $z_{(h)}$ via the so-called Mann iteration [Ber07, Chapter 4]. The works [GPC16, Gra17] refer to $\mathrm{argmin}_{x^i \in \mathcal{X}^i} J^i(x^i, z_{(h)})$ as optimal response rather than best response.

Algorithm 3 is a two-level algorithm, because at every iteration the agents compute

---

Algorithm 3: best response for Wardrop equilibrium

Initialization: $\tau > 0, k = 0, x^i_{(0)} \in \mathcal{X}^i \;\forall\, i, \lambda_{(0)} \in \mathbb{R}^m_{\geq 0}$.

Iterate

1. *Strategies are updated to a Wardrop equilibrium of* $\mathcal{G}_{\lambda_{(k)}}$

Initialization	$h = 0, \tilde{x}^i_{(0)} = x^i_{(k)}, z_{(0)} \in \mathbb{R}^n$
Iterate until convergence	$\tilde{x}^i_{(h+1)} = \underset{x^i \in \mathcal{X}^i}{\operatorname{argmin}} \; J^i(x^i, z_{(h)}) + \lambda^\top_{(k)} A_i x^i, \forall i$
	$z_{(h+1)} = \left(1 - \frac{1}{h}\right) z_{(h)} + \frac{1}{h}\left(\frac{1}{M}\sum_{j=1}^{M} \tilde{x}^j_{(h+1)}\right)$
	$h \leftarrow h + 1$
Upon convergence	$x_{(k+1)} = \tilde{x}_{(h)}$

2. *Dual variables are updated*

$$\lambda_{(k+1)} = \underset{\mathbb{R}^m_{\geq 0}}{\operatorname{Proj}} \left[\lambda_{(k)} - \tau(b - Ax_{(k+1)})\right]$$
$$k \leftarrow k + 1$$

---

the Wardrop equilibrium of the game without coupling constraints $\mathcal{G}_{\lambda_{(k)}}$. This is a task that in principle requires an infinite amount of iterations of the inner loop. We stress the fact that the following convergence result of Theorem 3 holds in the ideal case where the inner loop achieves exact convergence to the Wardrop equilibrium of $\mathcal{G}_{\lambda_{(k)}}$. Establishing convergence under inexact computation of the Wardrop equilibrium of $\mathcal{G}_{\lambda_{(k)}}$ is subject of future work.

The following assumption is used in Theorem 3 to prove convergence of Algorithm 3.

Assumption 4. For all $i \in \{1, \dots, M\}$ and $\lambda \in \mathbb{R}^m_{\geq 0}$, the mapping $z \mapsto \underset{x^i \in \mathcal{X}^i}{\operatorname{argmin}} \; J^i(x^i, z) + \lambda^\top A_i x^i$ is single valued; moreover, it is nonexpansive (Definition 4) or strongly pseudo-contractive (Definition 6). □

The authors of [GPC16] consider a cost function of linear quadratic form (3.31) and in [GPC16, Theorem 2] show that if $Q \succ 0, C = C^\top \succ 0$ or if $Q \succ 0, Q - C^\top Q^{-1} C \succ 0$, then Assumption 4 is satisfied. Lemma 4 states that under any of these two conditions, i.e., if $Q \succ 0, C = C^\top \succ 0$ or if $Q \succ 0, Q - C^\top Q^{-1} C \succ 0$, then F_{W} is strongly monotone, which is one of the assumptions of the following Theorem 3.

Theorem 3. Suppose that the operator F_{W} in (3.11b) is strongly monotone on $\mathcal{X}$ with constant α, that Assumptions 1, 3, 4 hold, and that $\mathcal{X}^i$ is bounded for all $i \in \{1, \dots, M\}$.

Then for any $x_{(0)}$, $\lambda_{(0)}$, the sequence $(x_{(k)})_{k=0}^{\infty}$ generated by Algorithm 3 with $\tau < \frac{2\alpha}{\|A\|^2}$ converges to a variational Wardrop equilibrium of $\mathcal{G}$.

Proof. We split the proof into three parts. First we show convergence of the inner loop, then of the outer loop, finally we conclude with a continuity argument.

Inner loop. In [GPC16, Theorem 3 and Corollary 1], it is shown that under Assumption 4, for any $\lambda_{(k)} \in \mathbb{R}^m_{\geq 0}$ the sequence $(\tilde{x}_{(h)})_{h=1}^{\infty}$ converges to $x_{(k)}$ such that

$$x^i_{(k)} = \operatorname*{argmin}_{x^i \in \mathcal{X}^i} \; J^i\left(x^i, \frac{1}{M}\sum_{i=1}^{M} x^i_{(k)}\right) + \lambda_{(k)}^\top A_i x^i. \tag{4.8}$$

In [GPC16, Theorem 1] it is shown that $x_{(k)}$ is an ε-Nash equilibrium for the game $\mathcal{G}(\lambda_{(k)})$, with $\varepsilon = \mathcal{O}(\frac{1}{M})$, but indeed (4.8) coincides with Definition 10 of Wardrop equilibrium[1].

Outer loop. Convergence of the outer loop is based on [PSP10, Proposition 8], but therein it is proved convergence to a Nash equilibrium, whereas Algorithm 3 achieves a Wardrop equilibrium, hence we report the entire proof for completeness. For each $\lambda \in \mathbb{R}^m_{\geq 0}$ define $F_{\mathrm{W}}(x;\lambda) \coloneqq F_{\mathrm{W}}(x) + A^\top \lambda$. Such operator is strongly monotone in x on the set $\mathcal{Q}$ with the same constant α as $F_{\mathrm{W}}(x)$. By Proposition 15, $\mathcal{G}(\lambda)$ has a unique variational Wardrop equilibrium which we denote by $\bar{x}_{\mathrm{W}}(\lambda)$. The outer loop update can be written as

$$\lambda_{(k+1)} = \operatorname*{Proj}_{\mathbb{R}^m_{\geq 0}}[\lambda_{(k)} - \tau(b - A\bar{x}_{\mathrm{W}}(\lambda_{(k)}))],$$

which is the update step of the projection algorithm reported in Algorithm 1 when applied to $\mathrm{VI}(\mathbb{R}^m_{\geq 0}, \Phi)$, with $\Phi(\lambda) \coloneqq b - A\bar{x}_{\mathrm{W}}(\lambda)$.

We now show cocoercitivity of Φ. To this end, consider $\lambda_1, \lambda_2 \in \mathbb{R}^m_{\geq 0}$ and the corresponding unique solutions $x_1 \coloneqq \bar{x}_{\mathrm{W}}(\lambda_1)$ of $\mathrm{VI}(\mathcal{X}, F_{\mathrm{W}}(x) + A^\top\lambda_1)$ and $x_2 \coloneqq \bar{x}_{\mathrm{W}}(\lambda_2)$ of $\mathrm{VI}(\mathcal{X}, F_{\mathrm{W}}(x) + A^\top\lambda_2)$. By Definition 1 of variational inequality,

$$(x_2 - x_1)^\top (F_{\mathrm{W}}(x_1) + A^\top \lambda_1) \geq 0\,, \tag{4.9a}$$
$$(x_1 - x_2)^\top (F_{\mathrm{W}}(x_2) + A^\top \lambda_2) \geq 0\,. \tag{4.9b}$$

Adding (4.9a) and (4.9b) we obtain $(x_2 - x_1)^\top (F_{\mathrm{W}}(x_1) - F_{\mathrm{W}}(x_2) + A^\top(\lambda_1 - \lambda_2)) \geq 0$, i.e., $(x_2 - x_1)^\top A^\top (\lambda_1 - \lambda_2) \geq (x_2 - x_1)^\top (F_{\mathrm{W}}(x_2) - F_{\mathrm{W}}(x_1))$. Since F_{W} is strongly monotone, it follows from the last inequality that

$$(Ax_2 - Ax_1)^\top (\lambda_1 - \lambda_2) \geq \alpha \|x_2 - x_1\|^2\,. \tag{4.10}$$

Moreover, since by definition of induced matrix norm $\|A(x_2 - x_1)\| \leq \|A\|\|x_2 - x_1\|$, then

$$\|x_2 - x_1\|^2 \geq \frac{\|A(x_2 - x_1)\|^2}{\|A\|^2}\,. \tag{4.11}$$

[1]This is consistent with the fact that it is an ε-Nash with $\varepsilon = \mathcal{O}(\frac{1}{M})$ thanks to Proposition 16.

Combining (4.10), (4.11), and adding and subtracting b, we obtain

$$(b - Ax_2 - (b - Ax_1))^\top(\lambda_2 - \lambda_1) \geq \frac{\alpha}{\|A\|^2}\|b - Ax_2 - (b - Ax_1)\|^2,$$

hence Φ is cocoercive in λ with constant $c_\Phi = \alpha/\|A\|^2$. Proposition 10 then guarantees that $\lambda_{(k)}$ converges to a solution of $\mathrm{VI}(\mathbb{R}^m_{\geq 0}, \Phi)$, provided that $\mathrm{VI}(\mathbb{R}^m_{\geq 0}, \Phi)$ admits at least a solution; the latter fact can be shown using the equivalence between 1) and 2) in Proposition 17 and the existence of a solution of $\mathrm{VI}(Q, F)$ by Proposition 15.

We have thus shown that $\lambda_{(k)}$ converges to a solution $\bar\lambda$ of $\mathrm{VI}(\mathbb{R}^m_{\geq 0}, \Phi)$. We now state that $\bar\lambda$ solves $\mathrm{VI}(\mathbb{R}^m_{\geq 0}, \Phi)$ if and only if $\mathbb{0}_m \leq \bar\lambda \perp (b - A\bar{x}_{\mathrm{W}}(\bar\lambda)) \geq \mathbb{0}_m$, which can be shown by writing down the KKT system (2.5) relative to $\mathrm{VI}(\mathbb{R}^m_{\geq 0}, \Phi)$, see [FP03, Proposition 1.1.3]. Hence we can conclude that $\lambda_{(k)} \to \bar\lambda$ and $\mathbb{0}_m \leq \bar\lambda \perp (b - A\bar{x}_{\mathrm{W}}(\bar\lambda)) \geq \mathbb{0}_m$.

Continuity argument. By the inner loop argument $x_{(k)} = \bar{x}_{\mathrm{W}}(\lambda_{(k)})$ is a Wardrop equilibrium of $\mathcal{G}(\lambda_{(k)})$ and by the outer loop argument $\lambda_{(k)}$ converges to a desired $\bar\lambda$. Moreover, [Nag90, Theorem 1.14] in our setup reads

$$\|\bar{x}(\bar\lambda) - \bar{x}_{\mathrm{W}}(\lambda_{(k)})\| \leq \frac{\|A\|}{\alpha}\|\bar\lambda - \lambda_{(k)}\|,$$

hence $x_{(k)} = \bar{x}_{\mathrm{W}}(\lambda_{(k)}) \to \bar{x}(\bar\lambda)$ and the proof is concluded. □

Remark 2 (Convergence rate)**.** We cannot provide general convergence rates for Algorithm 3, but if in Assumption 4 the best-response operator is strongly pseudocontractive, it is possible to replace $z_{(h+1)} = (1 - \frac{1}{h})z_{(h)} + \frac{1}{h}(\frac{1}{M}\sum_{j=1}^M \tilde{x}^j_{(h+1)})$ with $z_{(h+1)} = (1 - \frac{1}{\mu})z_{(h)} + \frac{1}{\mu}(\frac{1}{M}\sum_{j=1}^M \tilde{x}^j_{(h+1)})$ and guarantee geometric convergence of the inner loop for $\mu \in (0, 1]$ small enough, see e.g. [Ber07, Theorem 3.6 (iii)]. □

Comparison with the literature

Finding the equilibrium of a game modified by a dual variable which is in turn updated based on the coupling constraint violation is not a new idea; the principle is borrowed from dual ascent methods in convex optimization [BPC11, eq. (2.2)] and in the context of games the algorithm appears for Nash equilibrium in [PSP10, Proposition 8]. Our contribution consists in applying such idea to the Wardrop equilibrium problem and in using the existing best-response algorithm in [GPC16, Algorithm 1] in the inner loop.

At the end of Section 3.1 we explained that, to the best of our knowledge, the only other work defining a Wardrop equilibrium in terms of x is [Gra17], which also considers coupling constraints and proposes a single-level best-response algorithm for Wardrop equilibrium, which to date constitutes the only alternative to Algorithm 3. Thus we here perform a detailed comparison with the work [Gra17].

- The work [Gra17] focuses on a Wardrop equilibrium that satisfies the coupling constraints, rather than on the more restrictive concept of (generalized) Wardrop equilibrium of Definition 10. Mathematically, this amounts to looking for a solution of a relaxed version of (3.17), where (3.17b) is replaced by $g(x) \leq \mathbb{0}_m$, thus ignoring the orthogonality condition and the non-negativity of the dual variable. Whether one or the other equilibrium concept is more suitable depends on the application.

- The single-level algorithm in [Gra17] is very similar to the inner loop of Algorithm 3, the main difference being that z is updated in a more sophisticated manner which is needed to ensure convergence to an equilibrium satisfying the coupling constraints. Since it is single-level rather than two-level, it is in principle more attractive than Algorithm 3; moreover, [Gra17] proves a logarithmic convergence rate whereas we do not study the convergence rate.

- Regarding the assumptions on the cost functions, [Gra17] does not assume that J^i is continuously differentiable, nor that F_{W} is strongly monotone, nor a regularity condition similar to Assumption 4, but on the other hand studies the specific cost function $J^i(x^i, \sigma(x)) = v^i(x^i) + (C\sigma(x)) + c)^\top x^i$, with v^i strongly convex and C symmetric. In this case F_{W} is a gradient-operator by Proposition 2, i.e., VI($\mathcal{X}$,F_{W}) is equivalent to an optimization program.

- Regarding the coupling constraint, [Gra17] imposes a generic convex constraint on the average, i.e. $\sigma(x) \in \mathcal{S}$, with the requirement that $\mathcal{S} \subseteq \frac{1}{M}\sum_{i=1}^M \mathcal{X}^i$, with the last inclusion not needed in our setup. On the other hand, we assume the coupling constraint to be affine in x and anyway Algorithm 3 can be carried out in a parallel fashion only if the coupling constraint is expressed on the average, as commented above.

4.2 Gradient step parallel algorithm for Nash and Wardrop equilibrium

By exploiting the equivalent reformulation of VI($\mathcal{Q}, F$) as the extended VI($\mathcal{Y}, T$) given in Proposition 17, we propose here an algorithm that can be used to achieve a Nash or to achieve a Wardrop equilibrium. Solving VI($\mathcal{Y}, T$) instead of VI($\mathcal{Q}, F$) allows the design of a parallel algorithm, because the set $\mathcal{Y}$ is the Cartesian product $\mathcal{X}^1 \times \ldots \mathcal{X}^M \times \mathbb{R}^m_{\geq 0}$, and thus the individual constraint sets $\mathcal{X}^i$ are decoupled.

Algorithm 4 finds a solution of VI($\mathcal{Y}, T$), where T is as in (4.3), with $F = F_{\text{N}}$, and hence achieves a Nash equilibrium. If the same algorithm is used with $F = F_{\text{W}}$ it achieves a Wardrop equilibrium.

Algorithm 4: gradient step for Nash equilibrium / for Wardrop equilibrium

Initialization $\tau > 0, k = 0, x^i_{(0)} \in \mathcal{X}^i \; \forall\, i, \lambda_{(0)} \in \mathbb{R}^m_{\geq 0}$

Iterate $\sigma_{(k)} = \frac{1}{M} \sum_{i=1}^M x^i_{(k)}$

$x^i_{(k+1)} = \operatorname*{Proj}_{\mathcal{X}^i} [x^i_{(k)} - \tau \nabla_{x^i} \{J^i(x^i_{(k)}, \sigma(x_{(k)})) + A_i^\top \lambda_{(k)}\}], \forall i$

$\lambda_{(k+1)} = \operatorname*{Proj}_{\mathbb{R}^m_{\geq 0}} [\lambda_{(k)} - \tau(b - 2Ax_{(k+1)} + Ax_{(k)})]$

$k \leftarrow k + 1$

Regarding the information exchange, at every primal iteration the central operator measures the average agents' strategy $\sigma(x)$ and broadcasts it to the agents. Then each agent computes his new strategy $x^i_{(k+1)}$ by taking a gradient step, based on his previous strategy $x^i_{(k)}$, the previous average $\sigma(x_{(k)})$ and the previous dual variable $\lambda_{(k)}$. Given the new coupling constraint violation, the central operator updates the price to $\lambda_{(k+1)}$ and broadcasts it to the agents. If the coupling constraint is expressed on the average strategy $\sigma(x)$, then the central operator only needs to measure $\sigma(x)$ to perform the dual update and in this case the entire algorithm is parallel.

Let us now compare Algorithm 3 and Algorithm 4. In both cases the cost of each agent is modified by the additive term $A_i^\top \lambda_{(k)} x^i$, with the goal of enforcing the coupling constraint. The first key difference between the two algorithms is that in Algorithm 3 at the primal update each agent chooses the current minimizer of his cost function, whereas in Algorithm 4 he moves a step in the projected direction of the current gradient. The dual update is similar in the two cases, with the only difference being that Algorithm 4 features an extra $2\tau A x_{(k+1)}$ term. The second key difference is that Algorithm 3 is composed by two-loops, while Algorithm 4 consists in a single loop. A numerical comparison of the performances of the two algorithms is conducted in Section 6.1 for the electric vehicles application.

Theorem 4. Let Assumptions 1 and 3 hold.

- Let F_{N} in (3.11a) be strongly monotone in $\mathcal{X}$ with constant α and Lipschitz in $\mathcal{X}$ with constant L_F. Set $\tau > 0$ such that

$$\tau < \frac{-L_F^2 + \sqrt{L_F^4 + 4\alpha^2 \|A\|^2}}{2\alpha \|A\|^2}. \tag{4.13}$$

  Then for any $x_{(0)}$, $\lambda_{(0)}$, the sequence $(x_{(k)})_{k=0}^\infty$ generated by Algorithm 4 converges to a variational Nash equilibrium of $\mathcal{G}$.

- Let F_{W} in (3.11b) be strongly monotone on $\mathcal{X}$ with constant α and Lipschitz on $\mathcal{X}$ with constant L_F, then Algorithm 4 with $\nabla_{x^i} J^i(x^i_{(k)}, z)_{|z=\sigma(x)}$ in place of

$\nabla_{x^i} J^i(x^i_{(k)}, \sigma(x_{(k)}))$ converges to a variational Wardrop equilibrium, if τ satisfies (5.12). $\square$

Remark 3 (Convergence rate)**.** By specializing the result in [MW95] we proved in [PGP16, Proposition 1] that if the operator F is not only monotone but also affine and the set $\mathcal{X}$ is a polyhedron then for τ small enough Algorithm 4 converges R-linearly, i.e., $\limsup_{k\to\infty}(\|y_{(k)} - \bar{y}\|)^{\frac{1}{k}} < 1$. $\square$

The proof of Theorem 4 is not reported here to avoid distracting the reader from the main contributions of this thesis. Indeed, Algorithm 4 is the standard asymmetric projection algorithm [FP03, Algorithm 12.5.1] applied to VI$(\mathcal{Y}, T)$, which is known to converge when F is strongly monotone. The technical novelty in the proof compared to previous works is described in [GPP17].

Comparison with the literature

We start by showing that monotonicity of F implies monotonicity of T, as

$$(T(x_1,\lambda_1) - T(x_2,\lambda_2))^\top \left(\begin{bmatrix} x_1 \\ \lambda_1 \end{bmatrix} - \begin{bmatrix} x_2 \\ \lambda_2 \end{bmatrix} \right) = \begin{bmatrix} F(x_1) - F(x_2) + A^\top(\lambda_1 - \lambda_2) \\ -A(x_1 - x_2) \end{bmatrix}^\top \begin{bmatrix} x_1 - x_2 \\ \lambda_1 - \lambda_2 \end{bmatrix}$$

$$= \underbrace{(F(x_1) - F(x_2))^\top (x_1 - x_2)}_{\geq 0 \text{ as } F \text{ is strongly monotone}} + \underbrace{(\lambda_1 - \lambda_2)^\top A(x_1 - x_2) - (x_1 - x_2)^\top A^\top (\lambda_1 - \lambda_2)}_{=0} \geq 0.$$

It is also straightforward to show that Lipschitzianity of F implies Lipschitzianity of T. As a consequence, any algorithm for monotone variational inequalities converges to the desired variational equilibrium. A treatment of the classic ones can be found in [FP03, Chapter 12]. Specifically, it is possible to use the extragradient algorithm reported in Algorithm 2, which however requires twice as many primal and dual updates per iteration compared to Algorithm 4. Conducting a thorough comparison of existing algorithms for monotone variational inequalities applied to Nash and Wardrop equilibrium problems is beyond the scope of this thesis.

There is though a final point to be made. We explained in the literature comparison at the end of Section 3.1 that the work [Gra17], which is the only other study on the Wardrop equilibrium in terms of the strategies x, proposes a best-response algorithm. As a consequence, while there are plenty of gradient-step algorithms for the Nash equilibrium [CLL14, Ros65, YSM11, YP17, KNS12, ZF16], to the best of our knowledge Algorithm 4 can be seen as the first gradient-step algorithm that coordinates the agents to a Wardrop equilibrium in terms of x. This is by no means a major contribution by itself; the novelty rather consists in defining the Wardrop in terms of x in the first place, with the variational inequality reformulation and the application of an existing algorithm being just natural subsequent steps.

CHAPTER 5

A distributed algorithm

While Chapter 4 introduces parallel algorithms for Nash and Wardrop equilibrium, this Chapter 5 proposes a distributed algorithm, where the agents can only communicate locally with their neighbors as in the scheme of Figure 4.1a. The algorithm is based on gradient step, not on best response, and it is formulated for Nash equilibrium, but as we point out in Subsection 5.3.1, all the treatment carries through for Wardrop equilibrium as well. We perform a specific comparison with other distributed algorithms at the end of Section 5.1.

As a side contribution, to prove convergence of the distributed algorithm, we introduce a novel result on convergence of parametric variational inequalities, which is applicable beyond the context of games. The result is reported in the standalone Section 5.3, where we give a detailed explanation of how the result contributes to the existing literature on convergence of parametric variational inequalities.

The chapter results can be generalized, as shown in Subsection 5.3.1, and they are applied in Section 6.3 to a Cournot game with transportation costs.

5.1 Gradient step distributed algorithm for Nash and Wardrop equilibrium

We start by introducing the following assumption.

Assumption 5. The coupling constraint $\mathcal{C}$ in (3.2) is of the form

$$x \in \mathcal{C} := \{x \in \mathbb{R}^{Mn} \mid \hat{A}\sigma(x) \leq \hat{b}\}, \tag{5.1}$$

with $\hat{A} \in \mathbb{R}^{m \times n}$, $\hat{b} \in \mathbb{R}^m$, for some $m \in \mathbb{N}$. □

Assumption 5 is more stringent than Assumption 3, in that the coupling constraint is not only required to be an affine inequality, but also to be expressed on the average strategy. The coupling constraint in (5.1) can model the fact that the usage level for a certain commodity cannot exceed a fixed capacity, as in [KH12] and in [Gra17].

5.1.1 Communication limitations

The main objective of the chapter is to coordinate the agents' strategies to a Nash equilibrium of $\mathcal{G}$ by using a distributed algorithm that only requires communications over a pre-specified (sparse) communication network. We model such network by its adjacency matrix $T \in [0,1]^{M \times M}$, where the element $T_{ij} \in [0,1]$ is the weight that agent i assigns to communications received from agent j, with $T_{ij} = 0$ representing no communication. For brevity, we refer to T as communication network, even though it is the adjacency matrix of the communication network. Agent j is an in-neighbor of i if $T_{ij} > 0$ and an out-neighbor if $T_{ji} > 0$. We denote the sets of in- and out-neighbors of agent i as $\mathcal{N}^i_{\text{in}}$ and $\mathcal{N}^i_{\text{out}}$, respectively. Algorithm 5, which we are about to introduce, requires each agent to communicate both with his in-neighbors and out-neighbors. We introduce the following assumption on the communication network T.

Assumption 6 (Communication network)**.** The communication matrix T is primitive and doubly stochastic. □

The standard definitions of primitive and doubly-stochastic can be found in [OSAFM07], where graph theoretical conditions guaranteeing Assumption 6 are also presented. Doubly stochastic means that $\mathbb{1}_n^\top T = \mathbb{1}_n^\top$ and $T\mathbb{1}_n = \mathbb{1}_n$, while primitive means that there exists $h \in \mathbb{N}$ such that $[T^h]_{ij} > 0$ for all i, j. Loosely speaking, Assumption 6 ensures that if the agents communicate a sufficiently large number of times over T, they are able to recover the average of the agents' strategies.

5.1.2 The algorithm and its convergence

To compute an almost Nash equilibrium in a distributed fashion, we propose the following Algorithm 5, where at iteration k each agent i updates four variables:

- his strategy $x^i_{(k)}$,
- a dual variable $\lambda^i_{(k)}$ relative to the coupling constraint $\mathcal{C}$,
- a local average of his in-neighbors' strategies $\sigma^i_{\nu,(k)}$,
- a local average of his out-neighbors' dual variables $\mu^i_{\nu,(k)}$.

To overcome the fact that the communication network is sparse we assume that to compute $\sigma^i_{\nu,(k)}$ and $\mu^i_{\nu,(k)}$ the agents communicate not once but multiple times over the network T. The number of communications per update is denoted by $\nu \in \mathbb{N}$ and is a tuning parameter of the algorithm. The primal and the dual variable are in turn updated by a gradient-like step that depends on a second tuning parameter $\tau > 0$. By inspecting the primal and dual update steps, it is possible to recognize the basic structure of the asymmetric projection algorithm, whose simpler parallel version constitutes Algorithm 4. We finally note that both tuning parameters ν and τ are decided a priori and do not change during the algorithm execution.

Regarding the information exchange, each agent communicates with his out-neighbors to compute $\mu^i_{\nu,(k)}$, and with his in-neighbors to compute $\sigma_{\nu,(k+1)}$; he then uses these two quantities to perform the primal and dual update. The algorithm is distributed in that each agent only needs to communicate with his in-neighbors and out-neighbors.

Algorithm 5: for Nash equilibrium

Initialization $\tau > 0, \nu \in \mathbb{N}, k = 0, x^i_{(0)} \in \mathcal{X}^i \;\forall\, i, \lambda^i_{(0)} \in \mathbb{R}^m_{\geq 0} \;\forall\, i, \sigma^i_{\nu,(0)} = x^i_{(0)}$

Iterate *Communication: Dual*

$\mu^i_{\nu,(k)} = \lambda^i_{(k)}, \forall\, i$
repeat ν times
$\mu^i_{\nu,(k)} = \sum_{j \in \mathcal{N}^i_{\text{out}}} T_{ji}\, \mu^j_{\nu,(k)}, \forall\, i$

Update: Primal

$F^i_{\nu,(k)} = \nabla_{z_1} J^i(x^i_{(k)}, \sigma^i_{\nu,(k)}) + [T^\nu]_{ii} \nabla_{z_2} J^i(x^i_{(k)}, \sigma^i_{\nu,(k)}), \forall\, i$
$x^i_{(k+1)} = \operatorname{Proj}_{\mathcal{X}^i} [x^i_{(k)} - \tau(F^i_{\nu,(k)} + \hat{A}^\top \mu^i_{\nu,(k)})], \forall\, i$

Communication: Primal

$\sigma^i_{\nu,(k+1)} = x^i_{(k+1)}, \forall\, i$
repeat ν times
$\sigma^i_{\nu,(k+1)} = \sum_{j \in \mathcal{N}^i_{\text{in}}} T_{ij} \sigma^j_{\nu,(k+1)}, \forall\, i$

Update: Dual

$\lambda^i_{(k+1)} = \operatorname{Proj}_{\mathbb{R}^m_{\geq 0}} [\lambda^i_{(k)} - \tau(\hat{b} - 2\hat{A}\sigma^i_{\nu,(k+1)} + \hat{A}\sigma^i_{\nu,(k)})], \forall\, i$

$k \leftarrow k + 1$

To prove convergence, we introduce the following additional assumptions.

Assumption 7 (Coupling constraints)**.** The matrix $\hat{A}$ and the vector $\hat{b}$ in (5.1) are such that the following implication holds.

$$\{\hat{A}^\top \hat{s} = \mathbb{0}_n, \quad \hat{b}^\top \hat{s} \leq 0, \quad \hat{s} \geq \mathbb{0}_m\} \quad \Rightarrow \quad \hat{s} = \mathbb{0}_m. \qquad \square$$

Example 3. In this example we show that the simple coupling constraint of the form $\underline{b} \leq \sigma(x) \leq \bar{b}$ satisfies Assumption 7, when $\underline{b} < \bar{b}$ component-wise. To express $\mathcal{C}$ in terms of (5.1), it must be $\hat{A}^\top = [I_n, -I_n]$ and $\hat{b}^\top = [\bar{b}, -\underline{b}]$. Then the first condition of Assumption 7 reads $[I_n, -I_n]\hat{s} = \mathbb{0}_n$, which is equivalent to $\hat{s}_j = \hat{s}_{j+n}$ for all $j = 1, \ldots, n$.

The second condition reads $[\bar{b}, -\underline{b}]\hat{s} = 0$, which is equivalent to

$$\sum_{j=1}^{n} \bar{b}_j \hat{s}_j - \underline{b}_j \hat{s}_{j+n} = 0 \Leftrightarrow \sum_{j=1}^{n} (\bar{b}_j - \underline{b}_j)\hat{s}_j = 0.$$

Since $s_j \geq 0$ by the third condition and $\bar{b}_j - \underline{b}_j > 0$ by assumption, it follows $s_j = 0$ for all j, hence Assumption 7 is met. We show in the application of Section 6.3 another example of coupling constraint satisfying Assumption 7. ☐

We introduce a last regularity assumption, before stating the convergence result for Algorithm 5.

Assumption 8. The set $\mathcal{X}$ is bounded, the cost function $J^i(z_1, z_2)$ is twice continuously differentiable in z_1, z_2 for all i, the operator $F = [\nabla_{x^i} J^i(x^i, \sigma(x))]_{i=1}^M$ is strongly monotone in $\mathcal{X}$. ☐

Theorem 5. Suppose that Assumptions 1, 5, 6, 7, 8 hold and that the set $\mathcal{X}$ satisfies Slater's constraint qualification. Then, for every precision $\varepsilon > 0$, there exists a minimum number of communications $\nu_\varepsilon > 0$ such that, for every $\nu > \nu_\varepsilon$ and for every initial condition $(x_{(0)}, \lambda_{(0)}) \in \mathcal{X} \times \mathbb{R}^{Mm}_{\geq 0}$, the sequence $(x_{(k)})_{k=1}^{\infty}$ produced by Algorithm 5 converges to an ε-Nash equilibrium of $\mathcal{G}$ for τ small enough. ☐

The proof of Theorem 5 is the subject of the next Section 5.2, where in Theorem 7 we also provide a precise bound on τ.

Comparison with the literature

Distributed algorithms for Nash equilibrium of aggregative games are already present in the literature [KNS12, CLL14, PGG15b, KNS16], but these works do not handle coupling constraints. In other words, they build on the core assumption that the strategy sets are decoupled. Subsection 5.2.1 highlights how the proof of Theorem 5 simplifies in the absence of coupling constraints.

To the best of our knowledge, the only distributed algorithm available in the literature for aggregative games with coupling constraints is [LYH16]. However such algorithm is only applicable if the coupling constraints can be expressed as the solution set of a system of linear equations [LYH16, eq. (5)]. This is a very restrictive assumption that prevents the applicability of the algorithm suggested in [LYH16] to many practical cases.

We finally note that our work has some affinity with the distributed algorithms suggested in [YP17, ZF16, YSM11] to compute a Nash equilibrium of generic games (i.e., games that do not necessarily feature the aggregative structure considered here) with coupling constraints. The term "distributed" in all these references, however, refers to the fact that any specific agent is only allowed to communicate with the agents that affect his cost function. In average aggregative games the cost function of each agent is influenced by the strategy of all the other agents, because it is affected by the average agents' strategy. Consequently, the schemes proposed in [ZF16, YSM11, YP17] can theoretically be applied to aggregative games, but they would require communications among all the agents, or the presence of a central coordinator, as it was the case in Chapter 4.

5.2 Proof of convergence

The proof of the statement of Theorem 5, which is the subject of this section, unfolds as follows.

1. We define the auxiliary game $\mathcal{G}_\nu$ parametric in the number of communications ν and study in Lemmas 6 and 7 convergence of the primitives of $\mathcal{G}_\nu$ to those of $\mathcal{G}$.
2. In Theorem 6 we prove that the variational Nash equilibrium of $\mathcal{G}_\nu$ is an ε_ν-Nash equilibrium of $\mathcal{G}$, with $\varepsilon_\nu \to 0$ as $\nu \to \infty$. To this end we exploit a novel result on parametric convergence of variational inequalities, which is derived in Section 5.3.
3. In Theorem 7 we show that Algorithm 5 converges to a variational Nash of $\mathcal{G}_\nu$.

Combining Theorem 6 and Theorem 7 proves Theorem 5.

The auxiliary game

We start by introducing the auxiliary game $\mathcal{G}_\nu$. In each iteration of Algorithm 5 the agents communicate ν times over T; mathematically this is equivalent to communicating once over a fictitious network with adjacency matrix T^ν. Based on T^ν, we introduce for each agent $i \in \{1, \dots, M\}$ the local average $\sigma_\nu^i(x)$, defined as

$$\sigma_\nu^i(x) \coloneqq \sum_{j=1}^{M} [T^\nu]_{ij} x^j.$$

We define $\mathcal{G}_\nu$ as a game with same constraints and cost functions as in $\mathcal{G}$ in (3.3), except for the fact that each agent reacts to the local average $\sigma_\nu^i(x)$ instead of the global average $\sigma(x) = \frac{1}{M}\sum_{j=1}^M x^j$. Specifically, upon defining

$$\mathcal{C}_\nu \coloneqq \{x \in \mathbb{R}^{Mn} \,|\, \hat{A}\sigma_\nu^j(x) \leq \hat{b},\ \forall j \in \{1, \dots, M\}\}$$

we formally introduce the multi-communication network aggregative game as

$$\mathcal{G}_\nu := \begin{cases} \text{agents :} & \{1,\dots,M\} \\ \text{cost of agent } i: & J^i(x^i,\sigma_\nu^i(x)) \\ \text{individual constraint :} & \mathcal{X}^i \\ \text{coupling constraint :} & \mathcal{C}_\nu. \end{cases}$$

The definition of Nash equilibrium for $\mathcal{G}_\nu$ is the analogous of Definition 9 for $\mathcal{G}$, so we do not report it here. To introduce the variational Nash equilibrium of $\mathcal{G}_\nu$, let us define the following quantities.

$$F_\nu(x) := [\nabla_{x^i} J^i(x^i,\sigma_\nu^i(x))]_{i=1}^M, \tag{5.3a}$$
$$Q_\nu := \{x \in \mathcal{X} | A_\nu x \le b\}, \tag{5.3b}$$
$$Q_\nu^i(x^{-i}) := \{x^i \in \mathcal{X}^i | A_\nu x \le b\}, \tag{5.3c}$$
$$A_\nu := T^\nu \otimes \hat{A}, \tag{5.3d}$$
$$b := \mathbb{1}_M \otimes \hat{b}. \tag{5.3e}$$

By the analogous of Proposition 14, if $J^i(x^i,\sigma_\nu^i(x))$ is convex in x^i for all $x^{-i} \in \mathcal{X}^{-i}$, every solution $\bar{x}_\nu$ of VI(Q_ν,F_ν) is a Nash equilibrium of $\mathcal{G}_\nu$, called variational Nash equilibrium.

We also recall the quantities corresponding to (5.3) relative to $\mathcal{G}$

$$F(x) := [\nabla_{x^i} J^i(x^i,\sigma(x))]_{i=1}^M, \tag{5.4a}$$
$$Q := \{x \in \mathcal{X} | Ax \le b\}, \tag{5.4b}$$
$$Q^i(x^{-i}) := \{x^i \in \mathcal{X}^i | Ax \le b\}, \tag{5.4c}$$
$$A := \frac{1}{M}\mathbb{1}_M \mathbb{1}_M{}^\top \otimes \hat{A}. \tag{5.4d}$$

Note that throughout this chapter we use F as in (5.4a) rather than F_N for ease of notation. In (5.4) the coupling constraint $\mathcal{C}$ is expressed in the redundant form $Ax \le b$ (consisting of M repetitions of the constraint $\hat{A}\sigma(x) \le \hat{b}$) to match the structure of $A_\nu x \le b$ in (5.3).

We conclude by stating convergence of the operator F_ν to F as ν tends to infinity.

Lemma 6. Let Assumptions 1, 5 and 6 hold and $\mathcal{X}$ be bounded. Then

$$\lim_{\nu\to\infty} T^\nu = \frac{1}{M}\mathbb{1}_M \mathbb{1}_M{}^\top, \qquad \lim_{\nu\to\infty} A_\nu = A, \qquad \lim_{\nu\to\infty} F_\nu(x) = F(x),$$

with the last being uniform in x. The operators F_ν are bounded on $\mathcal{X}$ uniformly in ν. □

Proof. The fact that T primitive and doubly stochastic implies $\lim_{\nu\to\infty} T^\nu = \frac{1}{M}\mathbb{1}_M \mathbb{1}_M{}^\top$ is a well-known result in consensus theory, whose proof can be found in [Bul18, Corollary

5.1]. Convergence of A_ν to A follows immediately from definitions (5.3d), (5.4d) and the properties of the Kronecker product. Note that

$$F(x) = [\nabla_{z_1} J^i(x^i, \sigma(x)) + \frac{1}{M} \nabla_{z_2} J^i(x^i, \sigma(x))]_{i=1}^M,$$
$$F_\nu(x) = [\nabla_{z_1} J^i(x^i, \sigma_\nu^i(x)) + [T^\nu]_{ii} \nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))]_{i=1}^M.$$

Uniform convergence of F_ν to F follows by continuity of $\nabla_{z_1} J^i(z_1, z_2)$ and $\nabla_{z_2} J^i(z_1, z_2)$ in z_1, z_2 for all i (ensured by Assumption 1), by $[T^\nu]_{ii} \to \frac{1}{M}$, since $T^\nu \to \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top$, and by $\sigma_\nu^i(x) \to \sigma(x)$ uniformly in x.

Finally, as $\|\nabla_{z_1} J^i(z_1, z_2)\|$ and $\|\nabla_{z_2} J^i(z_1, z_2)\|$ are continuous functions over the compact set $\mathcal{X}^i \times \mathrm{conv}\{\mathcal{X}^1, \dots, \mathcal{X}^M\}$ by Assumption 1, then there exists $\kappa > 0$ such that $\|\nabla_{z_1} J^i(z_1, z_2)\| < \kappa$ and $\|\nabla_{z_2} J^i(z_1, z_2)\| < \kappa$. Note that $[T^\nu]_{ii} \le 1$ for all $i \in \{1, \dots, M\}$ and for all $\nu > 0$, since T and thus T^ν are non-negative and doubly stochastic. Then for all $x \in \mathcal{X}$

$$\begin{aligned}
&\|F_\nu(x)\|^2 = \sum_{i=1}^M \|\nabla_{z_1} J^i(x^i, \sigma_\nu^i(x)) + [T^\nu]_{ii} \nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))\|^2 \\
&\le \sum_{i=1}^M \big(\|\nabla_{z_1} J^i(x^i, \sigma_\nu^i(x))\|^2 + 2\|\nabla_{z_1} J^i(x^i, \sigma_\nu^i(x))\|\|\nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))\| + \|\nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))\|^2\big) \\
&\le \sum_{i=1}^M (\kappa^2 + 2\kappa^2 + \kappa^2) = 4M\kappa^2,
\end{aligned}$$

which proves that F_ν is bounded, uniformly in ν. □

The next lemma provides a sufficient condition for F_ν to be strongly monotone, which is then used in Theorem 6.

Lemma 7. Under Assumptions 1, 6 and 8, there exists ν_{SMON} such that F_ν is strongly monotone for all $\nu > \nu_{\text{SMON}}$. □

Proof. We point out that the statement is not dissimilar in spirit from Lemma 5, but the proof is conducted differently. Observe that

$$\begin{aligned}
&(F_\nu(x) - F_\nu(y))^\top (x - y) \\
&= (F(x) - F(y))^\top (x - y) + (F_\nu(x) - F(x) - (F_\nu(y) - F(y)))^\top (x - y) \\
&\ge \alpha_F \|x - y\|^2 - \underbrace{\|F_\nu(x) - F(x) - (F_\nu(y) - F(y))\|}_{\le \ell_\nu \|x-y\|} \|x - y\| \\
&\ge (\alpha_F - \ell_\nu)\|x - y\|^2,
\end{aligned}$$

where ℓ_ν is the Lipschitz constant of $F_\nu - F$, which is Lipschitz continuous because J^i is twice continuously differentiable for all i and $\mathcal{X}$ is compact (as it is closed by

Assumption 1 and bounded by Assumption 8). Moreover, for the same reason and since $\sigma_\nu(x) \to \sigma(x)$ uniformly in x, then $\nabla_x(F_\nu(x) - F(x)) \to 0$ uniformly in x and hence $\ell_\nu \to 0$, thus concluding the proof. □

Auxiliary theorems

Theorem 6 (Convergence of $\bar{x}_\nu$ to $\bar{x}$)**.** Suppose that Assumptions 1, 5, 6, 7, 8 hold. Then

1. The game $\mathcal{G}$ has a unique variational Nash equilibrium $\bar{x}$ and for any $\nu > \nu_{\text{SMON}}$, $\mathcal{G}_\nu$ has a unique variational Nash equilibrium $\bar{x}_\nu$. Moreover,

$$\lim_{\nu\to\infty} \bar{x}_\nu = \bar{x}. \tag{5.5}$$

2. For every $\varepsilon > 0$, there exists a $\nu_\varepsilon > \nu_{\text{SMON}}$ such that, for every $\nu > \nu_\varepsilon$, the variational Nash equilibrium $\bar{x}_\nu$ of $\mathcal{G}_\nu$ is an ε-Nash equilibrium of $\mathcal{G}$. □

Proof. 1) Existence and uniqueness of $\bar{x}$ and $\bar{x}_\nu$ solutions to VI(Q,F) and VI(Q_ν,F_ν) respectively is guaranteed by Proposition 7, because the operator F is strongly monotone by assumption and the operator F_ν (for $\nu > \nu_{\text{SMON}}$) is strongly monotone by Lemma 7. By Proposition 8, strong monotonicity of F_ν implies $\nabla_x F_\nu(x) \succ \alpha I_{Mn}$ for all x, which in turn implies $\nabla_{x^i} F^i_\nu(x) \succ \alpha I_n$ for all x, hence $J^i(x^i, \sigma^i_\nu(x))$ is convex in x^i. Consequently, Proposition 14 guarantees that $\bar{x}$ and $\bar{x}_\nu$ are the unique variational Nash equilibria of $\mathcal{G}$ and $\mathcal{G}_\nu$, respectively. The limit (5.5) follows from Theorem 8 in Section 5.3, which is a general result on convergence of parametric variational inequalities. The theorem is based on Assumption 9 in Section 5.3.

To verify such assumption note that Lemma 6 implies $A_\nu \to A$ and, for each $x \in \mathcal{X}$, $F_\nu(x) \to F(x)$; moreover, $b_\nu = b$ for all ν. We are left with proving (5.14), which in our setup reads

$$\{A^\top s = \mathbb{0}_{Mn}, \quad b^\top s \le 0, \quad s \ge \mathbb{0}_{Mn}\} \quad \Rightarrow \quad s = \mathbb{0}_{Mn}. \tag{5.6}$$

To prove (5.6) take $s := [s^1; \dots; s^M] \in \mathbb{R}^{Mn}$ such that $A^\top s = \mathbb{0}_{Mn}$, $b^\top s \le 0$ and $s \ge \mathbb{0}_{Mn}$ and define $\hat{s} := \sum_{j=1}^M s^j \in \mathbb{R}^n$. Then

$$\begin{aligned}
A^\top s = \mathbb{0}_{Mn} \quad &\Leftrightarrow \quad \left(\frac{1}{M}\mathbb{1}_M \mathbb{1}_M^\top \otimes \hat{A}^\top\right) s = \mathbb{0}_{Mn} \quad \Rightarrow \quad \hat{A}^\top \hat{s} = \mathbb{0}_n, \\
b^\top s \le 0 \quad &\Leftrightarrow \quad (\mathbb{1}_M^\top \otimes \hat{b}^\top) s \le 0 \quad \Rightarrow \quad \hat{b}^\top \hat{s} \le 0, \\
s \ge \mathbb{0}_{Mn} \quad &\Rightarrow \quad \hat{s} \ge \mathbb{0}_n.
\end{aligned}$$

By Assumption 7, we conclude $\hat{s} = \mathbb{0}_n$. Since $s \ge \mathbb{0}_{Mn}$, it must be $s = \mathbb{0}_{Mn}$, thus proving (5.6).

2) We divide the proof of this statement into two parts: i) we prove that $\bar{x}_\nu \in Q$ for any $\nu > 0$, and ii) we prove that condition (3.5) is satisfied, as these are the two conditions of Definition 9 of ε-Nash.
i) Being $\bar{x}_\nu$ a Nash equilibrium for $\mathcal{G}_\nu$, $\bar{x}_\nu \in \mathcal{Q}_\nu$, hence $\bar{x}_\nu \in \mathcal{X}$ and $\hat{A}\sigma^i_\nu(\bar{x}_\nu) \leq \hat{b}$ for all i. By summing over all i and dividing by M, we obtain

$$\hat{A}\left(\frac{1}{M}\sum_{i=1}^{M}\sigma^i_\nu(\bar{x}_\nu)\right) \leq \hat{b}. \tag{5.7}$$

However,

$$\sum_{i=1}^{M}\sigma^i_\nu(\bar{x}_\nu) = \sum_{i=1}^{M}\sum_{j=1}^{M}[T^\nu]_{ij}\bar{x}^j_\nu = \sum_{j=1}^{M}\left(\sum_{i=1}^{M}[T^\nu]_{ij}\right)\bar{x}^j_\nu = \sum_{j=1}^{M}\bar{x}^j_\nu = M\sigma(\bar{x}_\nu), \tag{5.8}$$

where the second to last equality holds because, by Assumption 6, T is doubly stochastic and so is T^ν as a consequence. By substituting (5.8) into (5.7) we obtain $\hat{A}\sigma(\bar{x}_\nu) \leq \hat{b}$, thus $\bar{x}_\nu \in \mathcal{Q}$ for any ν.
ii) Since $\bar{x}_\nu$ is a Nash equilibrium for $\mathcal{G}_\nu$, for all $i \in \{1,\dots,M\}$ and for all $x^i_\nu \in \mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ it holds

$$J^i(\bar{x}^i_\nu, \sum_j [T^\nu]_{ij}\bar{x}^j_\nu) \leq J^i(x^i_\nu, [T^\nu]_{ii}x^i_\nu + \sum_{j\neq i}[T^\nu]_{ij}\bar{x}^j_\nu). \tag{5.9}$$

For all i, $J^i(z_1, z_2)$ is continuously differentiable on the compact set $\mathcal{X}$, hence Lipschitz, because defined on a compact set. Then there exists a common Lipschitz constant L such that for all i, all $z_1, z^a_1, z^b_1 \in \mathcal{X}^i$, and all $z_2, z^a_2, z^b_2 \in \mathrm{conv}(\mathcal{X}^1,\dots,\mathcal{X}^M)$

$$\begin{aligned}\|J^i(z^a_1, z_2) - J^i(z^b_1, z_2)\| &\leq L\|z^a_1 - z^b_1\|,\\ \|J^i(z_1, z^a_2) - J^i(z_1, z^b_2)\| &\leq L\|z^a_2 - z^b_2\|.\end{aligned}$$

Let $D := \max_{z\in\mathrm{conv}\{\mathcal{X}^1,\dots,\mathcal{X}^M\}}\{\|z\|\}$ and $\delta(\nu) := \|\frac{1}{M}\mathbb{1}_M\mathbb{1}_M{}^\top - T^\nu\|_\infty$. Set $\varepsilon_1 := \varepsilon/(4LD)$. By Lemma 6, there exists $\nu_1 > 0$ such that for all $\nu > \nu_1$, $\delta(\nu) < \varepsilon_1$. Moreover,

$$\begin{aligned}&J^i(\bar{x}^i_\nu, \sigma(\bar{x}_\nu)) = J^i(\bar{x}^i_\nu, \textstyle\sum_j \frac{1}{M}\bar{x}^j_\nu) \leq J^i(\bar{x}^i_\nu, \sum_j [T^\nu]_{ij}\bar{x}^j_\nu) + L\|\sum_j(1/M - [T^\nu]_{ij})\bar{x}^j_\nu\|\\ &\leq J^i(\bar{x}^i_\nu, \textstyle\sum_j [T^\nu]_{ij}\bar{x}^j_\nu) + L\sum_j |1/M - [T^\nu]_{ij}|\|\bar{x}^j_\nu\|\\ &\leq J^i(\bar{x}^i_\nu, \textstyle\sum_j [T^\nu]_{ij}\bar{x}^j_\nu) + LD\max_i\{\sum_j |1/M - [T^\nu]_{ij}|\} = J^i(\bar{x}^i_\nu, \sum_j [T^\nu]_{ij}\bar{x}^j_\nu) + LD\delta(\nu)\\ &\overset{(5.9)}{\leq} J^i(x^i_\nu, [T^\nu]_{ii}x^i_\nu + \textstyle\sum_{j\neq i}[T^\nu]_{ij}\bar{x}^j_\nu) + LD\varepsilon_1 \leq J^i(x^i_\nu, \frac{1}{M}x^i_\nu + \sum_{j\neq i}\frac{1}{M}\bar{x}^j_\nu) + 2LD\varepsilon_1\\ &= J^i(x^i_\nu, \tfrac{1}{M}x^i_\nu + \textstyle\sum_{j\neq i}\frac{1}{M}\bar{x}^j_\nu) + \frac{\varepsilon}{2},\end{aligned} \tag{5.10}$$

for all $x^i_\nu \in \mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$, for all $\nu > \nu_1$. The last inequality in (5.10) can be proven with a chain of inequalities similar to the previous ones in (5.10). Condition (5.10) implies that (3.5) holds for all $x^i_\nu \in \mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$, as we used the fact that $\bar{x}_\nu$ is a Nash equilibrium for $\mathcal{G}_\nu$. Since we however want to prove that $\bar{x}_\nu$ is an ε-Nash for $\mathcal{G}$, we need to prove

that (3.5) holds for all $x^i \in Q^i(\bar{x}_\nu^{-i})$.
To this end, take any such $x^i \in Q^i(\bar{x}_\nu^{-i})$. Set $\varepsilon_2 := \varepsilon/(2L + \frac{2L}{M})$. Since we showed in the first statement that $\bar{x}_\nu \to \bar{x}$, by Lemma 9 in Section 5.3, there exists[1] $\nu_2 > \nu_{\text{SMON}}$ such that for all $\nu > \nu_2$ and all $i \in \{1, \dots, M\}$ there exists $\tilde{x}_\nu^i \in \mathcal{Q}_\nu^i(\bar{x}_\nu^{-i})$ such that $\|x^i - \tilde{x}_\nu^i\| \leq \varepsilon_2$. From (5.10) we know that since $\tilde{x}_\nu^i \in \mathcal{Q}_\nu^i(\bar{x}_\nu^{-i})$ then

$$\begin{aligned} &J^i(\bar{x}_\nu^i, \sigma(\bar{x}_\nu)) \leq J^i(\tilde{x}_\nu^i, \tfrac{1}{M}\tilde{x}_\nu^i + \textstyle\sum_{j\neq i} \tfrac{1}{M}\bar{x}_\nu^j) + \tfrac{\varepsilon}{2} \\ &\leq J^i(x^i, \tfrac{1}{M}x^i + \textstyle\sum_{j\neq i} \tfrac{1}{M}\bar{x}_\nu^j) + (L + \tfrac{L}{M})\varepsilon_2 + \tfrac{\varepsilon}{2} = J^i(x^i, \tfrac{1}{M}x^i + \textstyle\sum_{j\neq i} \tfrac{1}{M}\bar{x}_\nu^j) + \varepsilon. \end{aligned} \tag{5.11}$$

As (5.11) holds for all $i \in \{1, \dots, M\}$ and for all $x^i \in Q^i(\bar{x}_\nu^{-i})$ and given part i), we have proven that $\bar{x}_\nu$ is an ε-Nash equilibrium for $\mathcal{G}$, for all $\nu > \nu_\varepsilon := \max\{\nu_1, \nu_2\}$. □

Theorem 7 (Convergence of Algorithm 5)**.** Let Assumptions 1 and 5 hold and the set $\mathcal{X}$ satisfy Slater's constraint qualification of Definition 2. Suppose that for the value of ν used in Algorithm 5 the operator F_ν in (5.3a) is strongly monotone with constant $\alpha_\nu > 0$ and Lipschitz with constant $L_\nu > 0$. Let the step-size τ satisfy

$$\tau < \frac{-L_\nu^2 + \sqrt{L_\nu^4 + 4\alpha_\nu^2 \|A_\nu\|^2}}{2\alpha_\nu \|A_\nu\|^2}. \tag{5.12}$$

Then for every initial condition $(x_{(0)}, \lambda_{(0)}) \in \mathcal{X} \times \mathbb{R}_{\geq 0}^{Mm}$ the sequence $(x_{(k)})_{k=1}^\infty$ produced by Algorithm 5 converges to the unique variational Nash equilibrium of $\mathcal{G}_\nu$. □

Proof. Let us define $x_{(k)} := [x_{(k)}^i]_{i=1}^M$, $\lambda_{(k)} := [\lambda_{(k)}^i]_{i=1}^M$, $\sigma_{\nu,(k)} := [\sigma_{\nu,(k)}^i]_{i=1}^M$, $\mu_{\nu,(k)} := [\mu_{\nu,(k)}^i]_{i=1}^M$. Then the communication steps are equivalent to

$$\begin{aligned} \sigma_{\nu,(k)} &= (T^\nu \otimes I_n)\, x_{(k)}, \\ \mu_{\nu,(k)} &= (T^\nu \otimes I_m)^\top \lambda_{(k)}. \end{aligned}$$

Consequently, the update steps can be rewritten as

$$\begin{aligned} x_{(k+1)}^i &= \operatorname*{Proj}_{\mathcal{X}^i} [x_{(k)}^i - \tau(F_{\nu,(k)}^i + \hat{A}^\top \sum_{j=1}^M [T^\nu]_{ji} \lambda_{(k)}^j)], \\ \lambda_{(k+1)}^i &= \operatorname*{Proj}_{\mathbb{R}_{\geq 0}^m} [\lambda_{(k)}^i - \tau(\hat{b} - 2\hat{A} \sum_{j=1}^M [T^\nu]_{ij} x_{(k+1)}^j + \hat{A} \sum_{j=1}^M [T^\nu]_{ij} x_{(k)}^j)], \end{aligned}$$

for all $i \in \{1, \dots, M\}$ or, in compact form,

$$\begin{aligned} x_{(k+1)} &= \operatorname*{Proj}_{\mathcal{X}} [x_{(k)} - \tau\left(F_\nu(x_{(k)}) + A_\nu{}^\top \lambda_{(k)}\right)], \\ \lambda_{(k+1)} &= \operatorname*{Proj}_{\mathbb{R}_{\geq 0}^{Mm}} [\lambda_{(k)} - \tau(b - 2A_\nu x_{(k+1)} + A_\nu x_{(k)})]. \end{aligned} \tag{5.13}$$

[1]As Lemma 9 requires uniqueness of $\bar{x}_\nu$, we take $\nu_2 > \nu_{\text{SMON}}$. Note than ν_2 is independent from i.

The update (5.13) coincides with one iteration of the asymmetric projection algorithm given in [FP03, Algorithm 12.5.1] applied to VI($\mathcal{Q}_\nu, F_\nu$). Then [GPP17, Theorem 3] shows that, by choosing τ as in (5.12), which also implies $\tau < 1/\|A_\nu\|$, Algorithm 5 is guaranteed to converge to the unique solution of VI($\mathcal{Q}_\nu, F_\nu$), if Q_ν satisfies Slater's constraint qualification. This is the case because $\mathcal{X}$ satisfies Slater's constraint qualification and because the coupling constraint is affine, see [BV04, eq. (5.27)]. □

Remark 4. Note that $A_\nu = T^\nu \otimes \hat{A}$, hence $\|A_\nu\| = \|T^\nu\|\|\hat{A}\|$. Under Assumption 6, $\|T^\nu\| = 1$ since T is doubly stochastic. Hence in this case, one can use $\|\hat{A}\|$ instead of $\|A_\nu\|$ in (5.12). Moreover, Lemma 7 guarantees that strong monotonicity of F_ν assumed in Theorem 7 is met for $\nu > \nu_{\text{SMON}}$. □

5.2.1 Simplification in absence of coupling constraints

As anticipated in Section 5.1, distributed algorithms for the Nash of aggregative games without coupling constraints have already been proposed in the literature, for example in [KNS12, CLL14, PGG15b, KNS16]. We highlight here how the proofs of Theorems 6 and 7 and the steps of Algorithm 5 greatly simplify in the absence of coupling constraints (i.e. when $\mathcal{C} = \mathbb{R}^{Mn}$).

Regarding the first statement of Theorem 6, in the absence of coupling constraints the variational inequality relative to $\mathcal{G}_\nu$ and $\mathcal{G}$ feature the same set $\mathcal{X}$, which is not affected by the parameter ν. Convergence of $\bar{x}_\nu$ to $\bar{x}$ can thus be proven by using standard sensitivity analysis results for VI, as explained in details at the beginning of Section 5.3. In this case one can even prove Lipschitz continuity of the solution, so it is possible to derive bounds on the minimum number of communications ν needed to achieve any desired precision in (5.5).

Regarding the second statement of Theorem 6, in the absence of coupling constraints the fact that $\bar{x}_\nu$ is an ε-Nash equilibrium of $\mathcal{G}$ is a trivial consequence of (5.5) and of the fact that the cost functions are Lipschitz. The difficulty when introducing the coupling constraints are that i) the feasibility of $\bar{x}_\nu$ in $\mathcal{G}_\nu$ does not imply automatically feasibility of $\bar{x}_\nu$ in $\mathcal{G}$ and ii) in the definition of Nash equilibrium, the set of feasible deviations $\mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ in $\mathcal{G}_\nu$ is different from the set of feasible deviations $\mathcal{Q}^i(\bar{x}^{-i}_\nu)$ in $\mathcal{G}$ (without coupling constraints both these sets would instead be simply $\mathcal{X}^i$). This is why to prove the second statement of Theorem 6 one needs to show Hausdorff convergence of $\mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ to $\mathcal{Q}^i(\bar{x}^{-i}_\nu)$ as $\nu \to \infty$, as done in Lemma 8 of Section 5.3.

Regarding Algorithm 5 and Theorem 7, in the absence of coupling constraints one needs to solve VI($\mathcal{X}, F_\nu$) in a distributed fashion. Since the constraint set $\mathcal{X}$ can be decoupled among the agents, the standard projection algorithm [FP03, Algorithm 12.1.1] is distributed and it is guaranteed to converge, because F_ν is strongly monotone. In other words, one can run Algorithm 5 performing only the primal steps, with simplified strategy update $x^i_{(k+1)} = \operatorname*{Proj}_{\mathcal{X}^i}[x^i_{(k)} - \tau F^i_{\nu,(k)}]$, which is distributed.

5.3 Novel convergence result for parametric variational inequalities

The notation used in this section is disjoint from the rest of the chapter, because we present a standalone result on convergence of parametric variational inequalities, which is at the core of the proof of Theorem 6. Specifically, we study the convergence of the solution $\bar{x}_\theta$ of $\mathrm{VI}(\mathcal{Q}_\theta, F_\theta)$ to the solution $\bar{x}_{\hat{\theta}}$ of $\mathrm{VI}(\mathcal{Q}_{\hat{\theta}}, F_{\hat{\theta}})$ when $\theta \to \hat{\theta}$ and both the set and the operator are affected by the parameter θ. In the literature on convergence of solutions of parametric variational inequalities it is common to assume that $F_{\hat{\theta}}$ is strongly monotone and that F_θ converges uniformly to $F_{\hat{\theta}}$ as $\theta \to \hat{\theta}$. Besides that, the literature on the topic can be divided into three classes, based on the assumptions on the sets.

1) The first class of results focuses on sets that do not change, so that only the operator is affected by the parameter, and studies convergence of the solution $\bar{x}_\theta$ of $\mathrm{VI}(\mathcal{Q}, F_\theta)$ to the solution $\bar{x}_{\hat{\theta}}$ of $\mathrm{VI}(\mathcal{Q}, F_{\hat{\theta}})$. If the set $\mathcal{Q}$ is closed and convex, F_θ is Lipschitz in θ uniformly in x and $F_{\hat{\theta}}$ is strongly monotone, then the solution is Lipschitz continuous [Nag13, Theorem 1.14], [FP03, Section 5.3]. Strong monotonicity of $F_{\hat{\theta}}$ can be relaxed if the set $\mathcal{Q}$ is a polyhedron [QM89].

2) The second class of results [Tob86, Kyp87, Daf88] focuses on parametric sets that can be described as $\mathcal{Q}_\theta \coloneqq \{x \in \mathbb{R}^n \,|\, g(x,\theta) \leq \mathbb{0}_m\}$ for a suitable parametric function $g(x,\theta)$. Assuming that $g(x,\theta)$ converges uniformly in x to $g(x,\hat{\theta})$ as $\theta \to \hat{\theta}$ and that at $\bar{x}_{\hat{\theta}}$ the linear independence constraint qualification [FP03, p. 253] holds, it can be shown that the parametric solution $\bar{x}_\theta$ is locally Lipschitz continuous around $\hat{\theta}$. Such results have been applied to games, as for example in [PR03, Tob90].

3) The third class of results [Daf88, Mos69] is the most general and only assumes that $\mathcal{Q}_\theta$ converges to $\mathcal{Q}_{\hat{\theta}}$ according to the Kuratowski set convergence definition. In this case one can prove continuity of $\bar{x}_\theta$ around $\hat{\theta}$. We are not aware of results proving local Lipschitz continuity in this case.

Here we do not assume the linear independence constraint qualification, because this is difficult to guarantee a priori for $\mathcal{G}$. Instead, we focus on a specific form of the sets and prove convergence in Kuratowski as well as Hausdorff distance, whose definitions are in Section 2.4. We then exploit the results in the third class described above, specifically those of [Daf88], [Mos69], to show continuity of the VI solution. It is important to highlight that we do not consider a continuous parameter θ tending to $\hat{\theta}$, but we rather focus on the slightly less general case of a discrete parameter $\nu \in \mathbb{N}$ that tends to infinity, for two different reasons. The first is that Theorem 6 requires a continuity result on a discrete parameter, while the second lays in the fact that we could not find the continuous counterpart of the discrete results on continuity of a sequence of sets provided by [Mos69, SW79], which are used to prove Lemma 8. Specifically, we consider sets $\mathcal{Q}_\nu$ and $\mathcal{Q}_\infty$ that take the form

$$\mathcal{Q}_\nu := \{x \in \mathcal{X} \subset \mathbb{R}^n | A_\nu x \leq b_\nu\},$$
$$\mathcal{Q}_\infty := \{x \in \mathcal{X} \subset \mathbb{R}^n | A_\infty x \leq b_\infty\},$$

with $\mathcal{X}$ convex and compact, $A_\nu, A_\infty \in \mathbb{R}^{m\times n}$, $b_\nu, b_\infty \in \mathbb{R}^m$, and consider operators $F_\nu : \mathcal{X} \to \mathbb{R}^n$, $F_\infty : \mathcal{X} \to \mathbb{R}^n$. Our result can also be interpreted as an extension of [BD95, Boo63] on parametric quadratic programs, and of [QM89, Yen95] on parametric variational inequalities over polyhedral sets, in that we consider sets that are obtained as the intersection of a parametric polyhedron with a generic convex and compact set $\mathcal{X}$. Finally, we note that the parameter appears also in the matrix A_ν and not only in b_ν, as in [Yen95, HRT07] or as in model-predictive control [BMD02, Theorem 4]. The following assumption summarizes the specifics of our setup.

Assumption 9. Suppose that
a) The set $\mathcal{X} \subset \mathbb{R}^n$ is convex, compact and has non-empty interior.
Moreover, $\lim_{\nu\to\infty} A_\nu = A_\infty$, $\lim_{\nu\to\infty} b_\nu = b_\infty$ and

$$\{A_\infty^\top s = \mathbb{0}_n, b_\infty^\top s \leq 0, s \geq \mathbb{0}_m\} \quad \Rightarrow \quad s = \mathbb{0}_m. \tag{5.14}$$

b) The operator $F_\infty : \mathcal{X} \to \mathbb{R}^n$ is continuous, strongly monotone and there exists $\nu_{\text{SMON}} > 0$ such that $F_\nu : \mathcal{X} \to \mathbb{R}^n$ is continuous, strongly monotone for each $\nu > \nu_{\text{SMON}}$. For each $x \in \mathcal{X}$, $\lim_{\nu\to\infty} F_\nu(x) = F_\infty(x)$. □

We note here that (5.14) is less restrictive than the assumption that A_∞ has full row rank (i.e., $A_\infty^\top s = \mathbb{0}_n \Rightarrow s = \mathbb{0}_m$), which is usually imposed to guarantee the linear independence constraint qualification a priori, see for example [Daf88, Remark 2.2].

The next Lemma 8 proves Kuratowski and Hausdorff convergence of the sets. Kuratowski convergence of $\mathcal{Q}_\nu \to \mathcal{Q}_\infty$ is a key part of the proof of the following Theorem 8. The fact that $d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \to 0$ is instead used in Lemma 9, which is in turn needed for the proof of Theorem 6.

Lemma 8. If Assumption 9a holds then as $\nu \to \infty$ we have

$$\mathcal{Q}_\nu \to \mathcal{Q}_\infty \quad \text{and} \quad d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \to 0. \qquad \square$$

Proof. We define $S_\nu := \{x \in \mathbb{R}^n | A_\nu x \leq b_\nu\}$, $S_\infty := \{x \in \mathbb{R}^n | A_\infty x \leq b_\infty\}$ and start by showing that $S_\nu \to S_\infty$, that is, we prove that (2.19) holds. To show $\limsup S_\nu \subseteq S_\infty$, consider an arbitrary $\hat{x} \in \limsup S_\nu$, a sequence $(\nu_k)_{k=1}^\infty$, $\nu_k \to \infty$ and points $x_{\nu_k} \in S_{\nu_k}$ such that $x_{\nu_k} \to \hat{x}$. Since $A_{\nu_k} x_{\nu_k} \leq b_{\nu_k}$ for all $k > 0$, passing to the limit as $\nu_k \to \infty$ we obtain $A_\infty x \leq b_\infty$, hence $\hat{x} \in S_\infty$.

Conversely, we show $S_\infty \subseteq \liminf S_\nu$. Consider an arbitrary $\hat{x} \in S_\infty$, to show that $\hat{x} \in \liminf S_\nu$ one needs to construct a sequence $(x_\nu)_{\nu=1}^\infty$ with $x_\nu \in S_\nu$ and such that $x_\nu \to \hat{x}$. To this end, define $\tilde{A} : [0,1] \to \mathbb{R}^{m\times n}$, $\tilde{b} : [0,1] \to \mathbb{R}^{m\times 1}$ by

$$\tilde{A}(t) := \begin{cases} A_{\lfloor 1/t \rfloor} & \text{if } t \in (0,1] \\ A_\infty & \text{if } t = 0 \end{cases}$$
$$\tilde{b}(t) := \begin{cases} b_{\lfloor 1/t \rfloor} & \text{if } t \in (0,1] \\ b_\infty & \text{if } t = 0. \end{cases}$$

Note that $A_\nu = \tilde{A}(1/\nu)$, $b_\nu = \tilde{b}(1/\nu)$ for all $\nu \in \mathbb{N}$. By Assumption 9a,

$$\tilde{A}(t) \xrightarrow[t\to 0]{} \tilde{A}(0) = A_\infty,$$
$$\tilde{b}(t) \xrightarrow[t\to 0]{} \tilde{b}(0) = b_\infty.$$

Let

$$\tilde{x}(t) := \operatorname*{Proj}_{\tilde{S}(t)}[\hat{x}] = \begin{aligned} \operatorname*{argmin}_{x\in\mathbb{R}^n} \quad & \|x-\hat{x}\|^2 \\ \text{s.t.} \quad & \tilde{A}(t)x \le \tilde{b}(t) \end{aligned}$$

be the projection of $\hat{x}$ onto $\tilde{S}(t) := \{x \in \mathbb{R}^n | \tilde{A}(t)x \le \tilde{b}(t)\}$. Assumption 9a implies that the regularity conditions required by [BD95, Theorem 2.2] are met, hence $\tilde{x}(t)$ is continuous at $t = 0$, that is

$$\tilde{x}(t) \to \tilde{x}(0) = \operatorname*{Proj}_{\tilde{S}(0)}[\hat{x}] = \operatorname*{Proj}_{S_\infty}[\hat{x}] = \hat{x}\,.$$

Consider now the sequence $(x_\nu := \tilde{x}(1/\nu))_{\nu=1}^\infty$. Clearly $x_\nu \in \tilde{S}(1/\nu) = S_\nu$ and we have $\lim_{\nu\to\infty} x_\nu = \hat{x}$, thus proving that $\hat{x} \in \liminf S_\nu$. We have thus shown that $S_\nu \to S_\infty$. Since $\mathcal{Q}_\nu = S_\nu \cap \mathcal{X}$ and $\mathcal{Q}_\infty = S_\infty \cap \mathcal{X}$, $\mathcal{X}$ is closed and convex with non-empty interior and S_ν is closed and convex for all ν, by [Mos69, Lemma 1.4] we have that $\mathcal{Q}_\nu \to \mathcal{Q}_\infty$.

To conclude, since $\mathcal{Q}_\nu$ are closed subsets of $\mathbb{R}^n$ for all ν and $\mathcal{Q}_\infty$ is compact and non-empty, using [SW79, Theorem 3] we obtain that $\mathcal{Q}_\nu \to \mathcal{Q}_\infty$ implies $d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \to 0$, thus completing the proof. □

We use Lemma 8 and [Mos69, Theorem A(b)] to show that the solution of $\mathrm{VI}(\mathcal{Q}_\nu, F_\nu)$ converges to the solution of $\mathrm{VI}(\mathcal{Q}_\infty, F_\infty)$.

Theorem 8. If Assumption 9 holds then $\mathrm{VI}(\mathcal{Q}_\infty, F_\infty)$ has a unique solution $\bar{x}_\infty$ and, for $\nu > \nu_{\mathrm{SMON}}$, $\mathrm{VI}(\mathcal{Q}_\nu, F_\nu)$ has a unique solution $\bar{x}_\nu$. Moreover

$$\bar{x}_\nu \to \bar{x}_\infty.$$ □

Proof. The fact that $\mathrm{VI}(\mathcal{Q}_\infty, F_\infty)$ and $\mathrm{VI}(\mathcal{Q}_\nu, F_\nu)$ (for $\nu > \nu_{\mathrm{SMON}}$) have a unique solution is an immediate consequence of their strong monotonicity (Assumption 9b) and of Proposition 7. To prove convergence we apply Proposition 13 to the sequence $(\mathrm{VI}(\mathcal{Q}_\nu, F_\nu))_{\nu=1}^\infty$ and to $\mathrm{VI}(\mathcal{Q}_\infty, F_\infty)$. All the assumptions of Proposition 13 are direct consequences of Assumption 9, except for $\mathcal{Q}_\nu \to \mathcal{Q}_\infty$, which is proven in Lemma 8. Thus $\bar{x}_\nu \to \bar{x}_\infty$. □

To conclude the section, we state and prove Lemma 9, which is used in the proof of Theorem 6, but is reported here for ease of readability, as its proof needs Lemma 8. The notation of Lemma 9 is the same of Section 5.2.

Lemma 9. Let Assumptions 5, 7 hold, the set $\mathcal{X}$ be convex, compact, with non-empty interior, and consider the sequence $(\bar{x}_\nu \in \mathbb{R}^{Mn})_{\nu=1}^\infty$ with $\bar{x}_\nu \to \bar{x}$. Then, for every $\varepsilon > 0$ there exists $\tilde{\nu} > 0$ such that for all $\nu > \tilde{\nu}$, all $i \in \{1, \dots, M\}$ and all $x^i \in Q^i(\bar{x}_\nu^{-i})$ there exists an $\tilde{x}_\nu^i \in Q_\nu^i(\bar{x}_\nu^{-i})$ such that $\|x^i - \tilde{x}_\nu^i\| \le \varepsilon$, where $Q^i(\bar{x}_\nu^{-i})$ and $Q_\nu^i(\bar{x}_\nu^{-i})$ are defined in (5.4c) and (5.3c) respectively. □

Proof. We show this statement in two steps. Specifically, we show that for every $\varepsilon > 0$ there exists $\tilde{\nu} > 0$ such that for all $\nu > \tilde{\nu}$ and all $i \in \{1, \dots, M\}$

1. $d_H(Q^i(\bar{x}^{-i}), Q^i(\bar{x}_\nu^{-i})) \le \varepsilon/2$, and
2. $d_H(Q^i(\bar{x}^{-i}), \mathcal{Q}_\nu^i(\bar{x}_\nu^{-i})) \le \varepsilon/2$.

The conclusion then follows by the triangular inequality of the Hausdorff distance.
1) Note that

$$Q^i(\bar{x}^{-i}) := \{x^i \in \mathcal{X}^i | \hat{A}x^i \le M\hat{b} - \sum_{j \ne i} \hat{A}\bar{x}^j =: b^i\},$$
$$Q^i(\bar{x}_\nu^{-i}) := \{x^i \in \mathcal{X}^i | \hat{A}x^i \le M\hat{b} - \sum_{j \ne i} \hat{A}\bar{x}_\nu^j =: b_\nu^i\}.$$

By assumption $\bar{x}_\nu \to \bar{x}$. Consequently, $b_\nu^i \to b^i$. We now show that the implication

$$\{\hat{A}^\top \hat{s} = \mathbb{0}_n, \quad (b^i)^\top \hat{s} \le 0, \quad \hat{s} \ge \mathbb{0}_m\} \quad \Rightarrow \quad \hat{s} = \mathbb{0}_m$$

holds. The inequalities $(b^i)^\top \hat{s} \le 0$ and $\hat{A}^\top \hat{s} = \mathbb{0}_n$ imply

$$M\hat{b}^\top \hat{s} \le \Big(\sum_{j \ne i} \hat{A}\bar{x}^j\Big)^\top \hat{s} = \sum_{j \ne i} (\bar{x}^j)^\top (\hat{A}^\top \hat{s}) = 0 \Rightarrow \quad \hat{b}^\top \hat{s} \le 0.$$

By Assumption 7 we obtain $\hat{s} = \mathbb{0}_m$. Consequently, the sets $Q^i(\bar{x}^{-i}), Q^i(\bar{x}_\nu^{-i})$ satisfy (5.14) and hence Assumption 9a, so the conclusion follows by Lemma 8.
2) It can be proven similarly as the previous one. Note that the value of $\tilde{\nu}$ used in the proof might depend on $\hat{A}, \hat{b}, \bar{x}$, but not on $\bar{x}_\nu$. This is needed to apply Lemma 9 in Theorem 6 and it comes from proving the statement in two steps instead of applying Lemma 8 directly to $Q_\nu^i(\bar{x}_\nu^{-i}), Q^i(\bar{x}_\nu^{-i})$. □

5.3.1 Generalizations

In the following subsections, we briefly comment on some immediate generalizations of the results above, that were omitted to keep the exposition simple.

Distributed algorithm for network aggregative games

Algorithm 5 is used here to find an ε-Nash of $\mathcal{G}$. However, if we assume that F_ν in (5.3a) is strongly monotone when $\nu = 1$, then Algorithm 5 can be used to find the variational Nash equilibrium of any network aggregative game, as defined in [PGG15a] with network T, by setting $\nu = 1$. Algorithm 5 thus constitutes an alternative to the distributed algorithms derived for generic games in [YP17, ZF16]. Moreover, note that if we set $T = \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top$ and $\nu = 1$ then Algorithm 5 achieves the variational Nash equilibrium of $\mathcal{G}$, but communications among all agents or the presence of a central coordinator are required.

Weighted average

The above results can be immediately generalized to aggregative games that depend on a weighted average $\sigma(x) = \sum_{i=1}^M w_i x^i$, for some $w_i > 0$ instead of the average $\sigma(x) = \frac{1}{M}\sum_{i=1}^M x^i$ used above. We can impose $\sum_{i=1}^M w_i = 1$ without loss of generality. Then Assumption 6 can be modified to require T to be a primitive matrix with $w = [w_1, \dots, w_M] > 0$ as left eigenvector relative to the eigenvalue 1 (normalized such that $w^\top \mathbb{1}_M = 1$).

Local strategy sets of different dimensions

In the previous sections we have assumed that the strategy set of each agent has n components, i.e, $\mathcal{X}^i \subset \mathbb{R}^n$. Following the same arguments as in Section 6.3, our results can be generalized to the case in which each agent features a strategy set of different dimension, i.e, $\mathcal{X}^i \subset \mathbb{R}^{n_i}$, as in [Jen10, YP17] and the aggregate strategy is $\sigma^i(x) = \frac{1}{M}\sum_{j=1}^M [H^j x^j + h^j] \in \mathbb{R}^n$, for some matrices $H^j \in \mathbb{R}^{n\times n_j}$ and vectors $h^j \in \mathbb{R}^n$.

Wardrop instead of Nash equilibrium

The focus of this paper is on Nash equilibrium, but the setup and the results extend to the Wardrop equilibrium. If in the primal update of Algorithm 5 we use $F^i_{\nu,(k)} = \nabla_{z_1} J^i(x^i_{(k)}, \sigma^i_{\nu,(k)})$ (neglecting thus the second summand) then Algorithm 5 converges to a Wardrop equilibrium.

Finally, as for all distributed communication schemes, it would be important to assess the algorithm performance in the presence of delay and packet loss.

CHAPTER 6

Applications

6.1 Charging of electric vehicles

Electric-vehicles (EV) are foreseen to significantly penetrate the market in the coming years [ZWH15], therefore coordinating their charging schedules can provide services beneficial to the grid operations [GTL13]. By assuming that the energy price depends on the aggregate consumption, the works [MCH13, GPC16, PKL16] formulate the EV charging problem as an aggregative game and propose parallel schemes based on optimal response or gradient step, in the absence of coupling constraints. The proposed schemes steer the agents to Nash [PKL16] or Wardrop [MCH13, GPC16] equilibria[1]. Compared to the existing literature, we use the sufficient condition of Theorem 1 to guarantee strong monotonicity of F_{N}, even for the case of $v^i = 0$ in (3.19), which is not handled by [MCH13, GPC16, PKL16, Gra17]. Moreover, we introduce constraints coupling the agents' charging profiles, which can model limits on the aggregate peak consumption (these are also introduced in the recent work [Gra17]). Finally, we study the distance between the aggregate strategies at the Nash and at the Wardrop equilibrium and we establish uniqueness of the dual variables associated to the violation of the coupling constraints.

Constraints

We consider M electric vehicles and we identify agent i with vehicle i. The state of charge of vehicle i at time t is described by the variable s_t^i. The time evolution of s_t^i is specified by the discrete-time system $s_{t+1}^i = s_t^i + b^i x_t^i$, $t = 1, \dots, n$, where x_t^i is the charging control (or energy consumption) over the time interval t and the parameter $b^i > 0$ is the charging efficiency. We assume that the charging control cannot take negative values and that at time t it cannot exceed $\tilde{x}_t^i \geq 0$. The final state of charge is constrained to $s_{n+1}^i \geq \eta^i$, where $\eta^i \geq 0$ is the desired state of charge of agent i. Denoting $x^i = [x_1^i, \dots, x_n^i]^\top \in \mathbb{R}^n$,

[1] As pointed out in Section 3.1, the works [MCH13, GPC16] do not recognize the limit point as Wardrop equilibrium, but rather conclude that it is an ε-Nash.

the individual constraint of agent i can be expressed as

$$x^i \in \mathcal{X}^i := \left\{ x^i \in \mathbb{R}^n \;\middle|\; \begin{array}{l} 0 \le x^i_t \le \tilde{x}^i_t, \quad \forall\, t = 1, \ldots, n \\ \sum_{t=1}^n x^i_t \ge \theta^i \end{array} \right\}, \tag{6.1}$$

where $\theta^i := (b^i)^{-1}(\eta^i - s^i_1)$, with $s^i_1 \ge 0$ the state of charge at the beginning of the time horizon. Besides the individual constraints $x^i \in \mathcal{X}^i$, we also introduce the coupling constraint

$$x \in \mathcal{C} := \left\{ x \in \mathbb{R}^{Mn} \;\middle|\; \frac{1}{M} \sum_{i=1}^{M} x^i_t \le K_t,\, \forall\, t = 1, \ldots, n \right\}, \tag{6.2}$$

indicating that at time t the grid cannot deliver more than $M \cdot K_t$ units of power to the vehicles. In compact form (6.2) reads as $(\mathbb{1}_M^\top \otimes I_n) x \le MK$, where $K := [K_1, \ldots, K_n]^\top$.

Cost function

The cost function of each agent represents his energy bill, which we model as

$$J^i(x^i, \sigma(x)) = \sum_{t=1}^{n} p_t \left(\frac{d_t + \sigma_t(x)}{\kappa_t} \right) x^i_t =: p(\sigma(x))^\top x^i. \tag{6.3}$$

The cost (6.3) can be interpreted as the sum over all intervals of the energy consumption x^i_t multiplied by the energy unit price p_t at that interval. In (6.3) we assume that the energy price for each time interval $p_t : \mathbb{R}_{\ge 0} \to \mathbb{R}_{>0}$ depends on the ratio between total consumption and total capacity $(d_t + \sigma_t(x))/\kappa_t$ at that time interval t. The quantity d_t is the non-EV demand at time t, which is inflexible and considered fixed within the game, divided by M, while $\sigma_t(x) := \frac{1}{M} \sum_{i=1}^M x^i_t$ is the EV demand at time t divided by M. The constant κ_t is the total production capacity divided by M as in [MCH13, eq. (6)]. The total production capacity κ_t is in general not related to the upper bound K_t.

We point out that the energy price p is a function of $\sigma(x)$ rather than $\sum_{i=1}^M x^i$, because we assume that the energy infrastructure scales with the number of agents, as argued in [MCH13, eq. (6),(7)]. In other words, agent i does not see his energy price increased by the mere fact that new customers join the market. The same reasoning underlies the choice of expressing the coupling constraint on $\sigma(x)$.

6.1.1 Theoretical guarantees

We define the game $\mathcal{G}_M^{\mathrm{EV}}$ as in (3.3), with $\mathcal{X}^i$, $\mathcal{C}$ and $J^i(x^i, \sigma(x))$ as in (6.1), (6.2) and (6.3) respectively. In the following corollary we cast the main results of Chapters 3 and 4 to the EV application.

Corollary 3. Consider a sequence of games $(\mathcal{G}_M^{\mathrm{EV}})_{M=1}^{\infty}$. Assume that there exists $\tilde{x}^0$ such that $\tilde{x}^i_t \le \tilde{x}^0$ for all $t \in \{1, \ldots, n\}, i \in \{1, \ldots, M\}$ and for each game $\mathcal{G}_M^{\mathrm{EV}}$. Moreover, assume that for each game $\mathcal{G}_M^{\mathrm{EV}}$ the set $\mathcal{Q} = \mathcal{C} \cap \mathcal{X}$ has non-empty interior and that for each t the price function p_t in (6.3) is twice continuously differentiable, strictly increasing and Lipschitz in $[0, \tilde{x}^0]$ with constant L_p. Moreover, assume

$$\min_{\substack{t \in \{1,\dots,n\} \\ z \in [0,\tilde{x}^0]}} \left(p_t'(z) - \frac{\tilde{x}^0 p_t''(z)}{8} \right) > 0. \tag{6.4}$$

Then:

1. A Wardrop and a Nash equilibrium exist for each game $\mathcal{G}_M^{\mathrm{EV}}$ of the sequence. Furthermore, every Wardrop equilibrium is an ε-Nash equilibrium with $\varepsilon = \frac{2n(\tilde{x}^0)^2 L_p}{M}$.

2. The function p is strongly monotone, hence for each game $\mathcal{G}_M^{\mathrm{EV}}$ there exists a unique $\bar{\sigma}$ such that $\sigma(\bar{x}_{\mathrm{W}}) = \bar{\sigma}$ for any variational Wardrop equilibrium $\bar{x}_{\mathrm{W}}$ of $\mathcal{G}_M^{\mathrm{EV}}$. Moreover for any variational Nash equilibrium $\bar{x}_{\mathrm{N}}$ of $\mathcal{G}_M^{\mathrm{EV}}$, $\|\sigma(\bar{x}_{\mathrm{N}}) - \sigma(\bar{x}_{\mathrm{W}})\| \leq \tilde{x}^0 \sqrt{\frac{2nL_p}{\alpha M}}$, where α is the monotonicity constant of p.

3. For each game $\mathcal{G}_M^{\mathrm{EV}}$ the operator F_{W} is monotone, hence the extragradient algorithm (Algorithm 2) with operator F_{W} converges to a variational Wardrop equilibrium of $\mathcal{G}_M^{\mathrm{EV}}$.

4. For each game $\mathcal{G}_M^{\mathrm{EV}}$ the operator F_{N} is strongly monotone. Hence, Algorithm 4 converges to the unique variational Nash equilibrium of $\mathcal{G}_M^{\mathrm{EV}}$. □

Proof. 1) We show that Assumption 1 holds. Indeed the sets $\mathcal{X}^i$ in (6.1) are convex and compact, the function g in (3.2) is affine and hence convex, and $\mathcal{Q}$ has non-empty interior by assumption. For each z fixed, the function $J^i(x^i, z)$ is linear hence convex in x^i. We prove in the last statement that F_{N} is strongly monotone. This is equivalent to $\nabla_x F_{\mathrm{N}}(x) \succ 0$ by Proposition 8, which by definition of $F_{\mathrm{N}}(x)$ implies $\nabla_{x^i}(\nabla_{x^i} J^i(x^i, \sigma(x))) \succ 0$, which implies convexity of $J^i(x^i, \sigma(x))$. Finally, $J^i(z_1, z_2)$ is continuously differentiable in $[z_1; z_2]$ because p_t is twice continuously differentiable. As Assumption 1 holds and $\mathcal{Q}$ is bounded, Proposition 15 guarantees the existence of a Nash and of a Wardrop equilibrium. The ε-Nash property is guaranteed by Proposition 16 upon verifying Assumption 2. This holds because $\cup_{i=1}^M \mathcal{X}^i \subseteq [0, \tilde{x}^0]^n = \mathcal{X}^0$ and because $J^i(z_1, z_2)$ is Lipschitz in z_2 in $\mathcal{X}^0$ with Lipschitz constant $L_2 = RL_p$, as by assumption L_p is the Lipschitz constant of p_t and (3.35) holds. Moreover, $R \coloneqq \max_{y \in \mathcal{X}^0}\{\|y\|\} = \|\tilde{x}^0 \mathbb{1}_n\| = \tilde{x}^0 \sqrt{\sum_{i=1}^n 1} = \tilde{x}^0 \sqrt{n}$. Then the expression $\varepsilon = 2RL_2/M$ given in Proposition 16 in this case reads $\varepsilon = (2n(\tilde{x}^0)^2 L_p)/M$.
2) The fact that each p_t is strictly increasing in $[0, \tilde{x}^0]$ implies that $\nabla_z p(z) \succ 0$ in $[0, \tilde{x}^0]^n$, where $p(z) \coloneqq \left[p_1(\frac{d_1 + z_1}{\kappa}), \dots, p_n(\frac{d_n + z_n}{\kappa})\right]^\top$. In turn $\nabla_z p(z) \succ 0$ guarantees strong monotonicity of p in the compact set $[0, \tilde{x}^0]^n$ by Proposition 8. This, together with Assumptions 1 and 2 verified above, allows us to use the bound (3.38) in Theorem 2, which

proves the statement.
3) Since $\mathcal{X}$ is closed and convex, Proposition 12 guarantees that the extragradient algorithm converges to a Wardrop equilibrium, because F_{W} is monotone (by the first statement of Lemma 2) and Lipschitz (as J^i is twice continuously differentiable on a compact set), and a variational Wardrop equilibrium exists by Proposition 15.
4) Assumption 1, which has been shown to hold in the first statement, and Assumption 3, which trivially holds, allow us to use Theorem 4, because F_{N} is strongly monotone by Theorem 1, as condition (6.4) is identical to (3.23). □

6.1.2 Uniqueness of dual variables.

Corollary 3 shows that under condition (6.4) the operator F_{N} of $\mathcal{G}_M^{\mathrm{EV}}$ is strongly monotone, hence the game $\mathcal{G}_M^{\mathrm{EV}}$ admits a unique variational Nash equilibrium by Proposition 15. We study here the uniqueness of the associated dual variables $\bar{\lambda}_{\mathrm{N}}$ introduced in Proposition 17. Guaranteeing unique dual variables might be important to convince the vehicle owners to participate in the proposed scheme, as the dual variables represent the penalty price associated to the coupling constraint. Define $R^{\text{tight}} \subseteq \{1, \dots, n\}$ as the set of instants in which the coupling constraint $\mathcal{C}$ is active. We provide a sufficient condition for uniqueness of the dual variables which relies on a slight modification of the linear-independence constraint qualification.

Lemma 10. Assume that $\mathcal{Q} = \mathcal{C} \cap \mathcal{X}$ has non-empty interior, that for each t the price function p_t in (6.3) is continuously differentiable, strictly increasing. Let condition (6.4) hold and consider the unique variational Nash equilibrium $\bar{x}_{\mathrm{N}}$ of $\mathcal{G}_M^{\mathrm{EV}}$. If there exists a vehicle i such that

- $\bar{x}^i_{\mathrm{N},t} \notin \{0, \tilde{x}^i_t\}$ for all $t \in R^{\text{tight}}$ and
- $\bar{x}^i_{\mathrm{N},t'} \notin \{0, \tilde{x}^i_{t'}\}$ for some $t' \notin R^{\text{tight}}$,

then the dual variables $\bar{\lambda}_{\mathrm{N}}$ associated to the coupling constraint (6.2) are unique. □

Before reporting the proof, we note that the sufficient condition of Lemma 10 is to be verified a-posteriori; in other words, it depends on the primal solution $\bar{x}_{\mathrm{N}}$. In the numerical analysis presented in the following such sufficient condition always holds. Uniqueness of the dual variables associated to the coupling constraint of an aggregative game has been studied also in [YSM11, Theorem 4], where the conditions in the bullets of Lemma 10 are not required but p is restricted to be affine. A sufficient and necessary condition for uniqueness of the dual variables is the strict Mangasarian-Fromovitz constraint qualification [Kyp85], but this requires an a-posteriori check on both the primal variable $\bar{x}_{\mathrm{N}}$ and on the dual variables $\bar{\lambda}_{\mathrm{N}}$.

Proof. Existence of the dual variables relative to VI(Q, F_N) follows from Proposition 17, so we are left with proving uniqueness. The constraint set Q, which is the intersection of (6.1) and (6.2), can be expressed as $\Gamma x \leq \gamma$ with

$$\Gamma = \begin{bmatrix} I_{M\cdot n} \\ -I_{M\cdot n} \\ -I_M \otimes \mathbb{1}_n^\top \\ \mathbb{1}_M^\top \otimes I_n \end{bmatrix}, \quad \gamma = \begin{bmatrix} \tilde{x} \\ \mathbb{0}_{Mn} \\ -\theta \\ MK \end{bmatrix},$$

where $\theta = [\theta^1, \dots, \theta^M]^\top$, and $\tilde{x} = [[\tilde{x}_t^i]_{t=1}^n]_{i=1}^M$. Let us partition the constraint matrix Γ into its individual part Γ_1 and coupling part Γ_2

$$\Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix}, \; \Gamma_1 = \begin{bmatrix} I_{M\cdot n} \\ -I_{M\cdot n} \\ -I_M \otimes \mathbb{1}_n^\top \end{bmatrix}, \; \Gamma_2 = \begin{bmatrix} \mathbb{1}_M^\top \otimes I_n \end{bmatrix} \tag{6.5}$$

and $\gamma = [\gamma_1^\top, \gamma_2^\top]^\top$ accordingly. By Proposition 4, the KKT conditions for VI($\mathcal{Q}, F_\mathrm{N}$) at the primal solution $\bar{x}_\mathrm{N}$ are

$$F_\mathrm{N}(\bar{x}_\mathrm{N}) + \Gamma_1^\top \mu + \Gamma_2^\top \lambda = \mathbb{0}_{Mn}, \tag{6.6a}$$

$$\mathbb{0}_{2Mn} \leq \mu \perp \gamma_1 - \Gamma_1 \bar{x}_\mathrm{N} \geq \mathbb{0}_{2Mn}, \tag{6.6b}$$

$$\mathbb{0}_M \leq \lambda \perp \gamma_2 - \Gamma_2 \bar{x}_\mathrm{N} \geq \mathbb{0}_M. \tag{6.6c}$$

Define $\tilde{\mu}$ and $\tilde{\lambda}$ as the dual variables corresponding to the active constraints (the other dual variables must be zero due to (6.6b) and (6.6c)). The KKT system (6.6) in $\tilde{\mu}, \tilde{\lambda}$ only reads

$$\begin{aligned} &\tilde{\Gamma}_1^\top \tilde{\mu} + \tilde{\Gamma}_2^\top \tilde{\lambda} = -F_\mathrm{N}(\bar{x}_\mathrm{N}), \\ &\tilde{\mu}, \tilde{\lambda} \geq \mathbb{0}\,, \end{aligned} \tag{6.7}$$

where $\tilde{\Gamma}_1, \tilde{\Gamma}_2$ contain the subset of rows of Γ_1, Γ_2 corresponding to active constraints. To conclude the proof we need to show that (6.7) has a unique solution $\tilde{\lambda}$. To this end we apply the subsequent Lemma 11. By analyzing the expression of Γ_1 and Γ_2 in (6.5), one could show that the negation of the assumption of Lemma 11 is equivalent to the existence of $R' \subseteq R^\mathrm{tight}$ such that for each vehicle i it holds $\bar{x}_{\mathrm{N},t}^i \in \{0, \tilde{x}_r^i\}$ for all $t \in R'$ or $\bar{x}_{\mathrm{N},t}^i \in \{0, \tilde{x}_t^i\}$ for all $t \in \{1, \dots, n\} \setminus R'$. Such R' cannot exist by assumption. □

Lemma 11. Consider $A_1 \in \mathbb{R}^{m \times n_1}$, $A_2 \in \mathbb{R}^{m \times n_2}$, $b \in \mathbb{R}^m$. If the implication $A_1 x_1 + A_2 x_2 = 0 \;\Rightarrow\; x_1 = 0$ holds for all $x_2 \in \mathbb{R}^{n_2}$, then the linear system of equations $A_1 x_1 + A_2 x_2 = b$ has at most one solution in x_1.

Proof. Assume $A\tilde{x} = b$ and $A\hat{x} = b$, then $A_1\tilde{x}_1 + A_2\tilde{x}_2 = b$ and $A_1\hat{x}_1 + A_2\hat{x}_2 = b$ imply $A_1(\hat{x}_1 - \tilde{x}_1) + A_2(\hat{x}_2 - \tilde{x}_2) = 0$, which by assumption implies $\hat{x}_1 = \tilde{x}_1$. □

6.1.3 Numerical analysis

The numerical study is conducted on a set of heterogeneous agents. We set the price function to $p_t(z_t) = 0.15\sqrt{(d_t + \sigma_t(x))/\kappa_t}$ as in [MCH13, eq.(25)], and $n = 24$. The agents differ in θ^i, randomly chosen according to $\mathcal{U}[0.5, 1.5]$; they also differ in $\tilde{x}_t^i$, which is taken such that the charge is allowed in a connected interval, with left and right endpoints uniformly randomly chosen: within the interval, $\tilde{x}_t^i$ is constant and randomly drawn for each agent, according to $\mathcal{U}[1, 5]$; outside this interval, $\tilde{x}_t^i = 0$. The demand d_t is taken as the typical (non-EV) base demand over a summer day in the United States [MCH13, Figure 1]; $\kappa_t = 12$ kW for all t, and the upper bound $K_t = 0.55$ kW is picked such that the coupling constraint (6.2) is active in the middle of the night. Note that with these choices all the assumptions of Corollary 3 are met. In particular, for the given choice of p condition (6.4) holds because $p_t''(z) < 0$ for all z and all t. Figure 6.1 presents the aggregate consumption at the Nash equilibrium found by Algorithm 4, with stopping criterion $\|(x_{(k+1)}, \lambda_{(k+1)}) - (x_{(k)}, \lambda_{(k)})\|_\infty \leq 10^{-4}$.

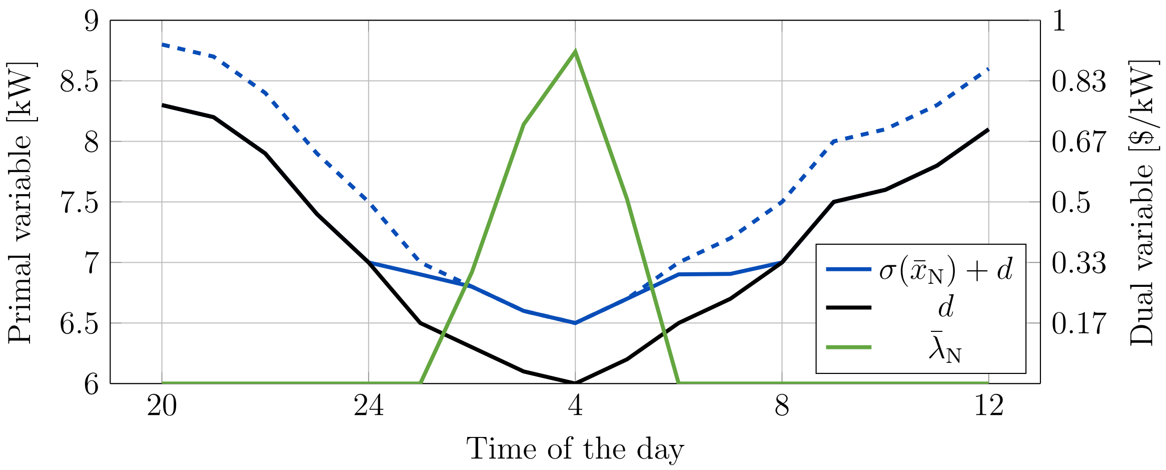


Figure 6.1: Aggregate EV demand $\sigma(\bar{x}_N)$ and dual variables $\bar{\lambda}_N$ for $M = 100$, subject to $\sigma(x) \leq 0.55$ kW. The region below the dashed line corresponds to $\sigma(x) + d \leq 0.55$ kW+d.

Figure 6.2 illustrates the bound $\|\sigma(\bar{x}_N) - \sigma(\bar{x}_W)\| \leq \tilde{x}^0\sqrt{\frac{2nL_p}{\alpha M}}$ of the second statement of Corollary 3. Since F_W is monotone but not strongly monotone, Algorithms 3 and 4 are not guaranteed to converge, hence the Wardrop equilibrium is computed with the extragradient algorithm (in Algorithm 2) with stopping criterion $\|(x_{(k+1)}, \lambda_{(k+1)}) - (x_{(k)}, \lambda_{(k)})\|_\infty \leq 10^{-4}$. The Nash $\bar{x}_N$ is computed instead with Algorithm 4, with the same stopping criterion. The ε-Nash property of the Wardrop equilibrium in the first statement of Corollary 3 can also be illustrated, resulting in a plot similar to Figure 6.2, which we omit here.

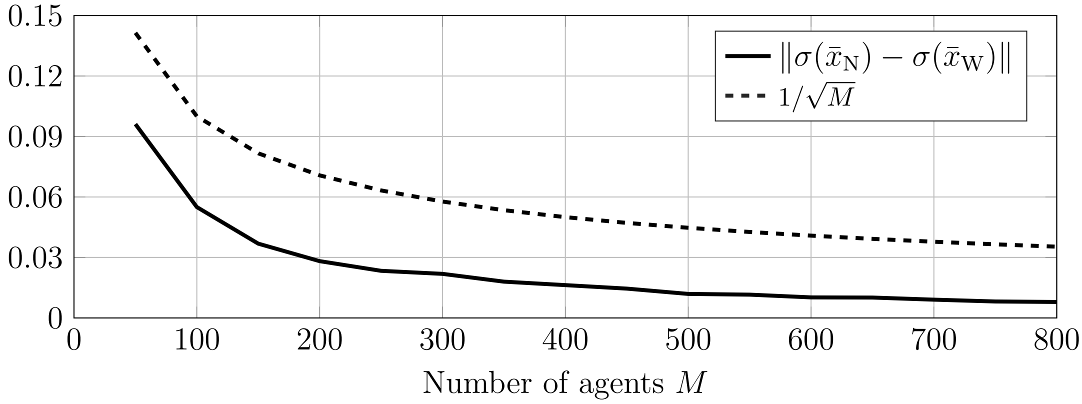


Figure 6.2: Distance between the aggregates $\sigma(\bar{x}_N)$ and $\sigma(\bar{x}_W)$ at the Nash and Wardrop equilibrium (solid line). Corollary 3 ensures that such distance is upper bounded by $\beta/\sqrt{M}$ for $\beta = \tilde{x}^0\sqrt{2nL_p/\alpha}$. The dotted line shows $1/\sqrt{M}$, illustrating that our bound has the right trend, while the constant $\beta \gg 1$ is, in this case, conservative.

The aggregate Wardrop strategy is valley filling

The first work [MCH13] that proposed to study the charging of EVs as a noncooperative game focused also on the valley filling property of the aggregate EV consumption, in a setup without coupling constraints (i.e., with $\mathcal{C} = \mathbb{R}^{Mn}$). The term valley filling refers to the fact that the overnight non-EV energy demand valley is filled by the EV energy consumption; the corresponding desirability for grid operations and social welfare is addressed in [MCH13, Lemma 3.1]. Specifically, according to [MCH13, eq. (9)] an aggregate strategy $\sigma(x)$ is valley filling if there exists a constant $\delta > 0$ such that

$$\sigma_t(x) > 0 \Rightarrow \sigma_t(x) + d_t = \delta. \tag{6.8}$$

This is indeed the case for $\sigma(\bar{x}_W)$, as visualized in Figure 6.3.

The reason for which $\sigma(\bar{x}_W)$ is valley filling can be understood by reinterpreting the EV game $\mathcal{G}_M^{\text{EV}}$ as the parallel road game of Example 2, with each time slot t corresponding to a parallel road that connects origin and destination, see Figure 3.1. The energy price p_t at a certain interval t corresponds to the travel time along the road t. Then (3.10) becomes

$$\begin{aligned}\sigma_t(\bar{x}_W) > 0 &\Rightarrow p_t(\sigma(\bar{x}_W)) = p_{\min}(\sigma(\bar{x}_W)),\\ \sigma_t(\bar{x}_W) = 0 &\Rightarrow p_t(\sigma(\bar{x}_W)) \geq p_{\min}(\sigma(\bar{x}_W)),\end{aligned}$$

which implies (6.8) with $\delta = p_{\min}(\sigma(\bar{x}_W))$, thus proving that $\sigma(\bar{x}_W)$ is valley filling. Other works [GPC16, VGA15] studied the valley filling property for EVs along with [MCH13], and in particular showed that the Nash strategies are almost valley filling, but to the best of our knowledge this is the first time that the aggregate Wardrop equilibrium is proved to be valley filling.

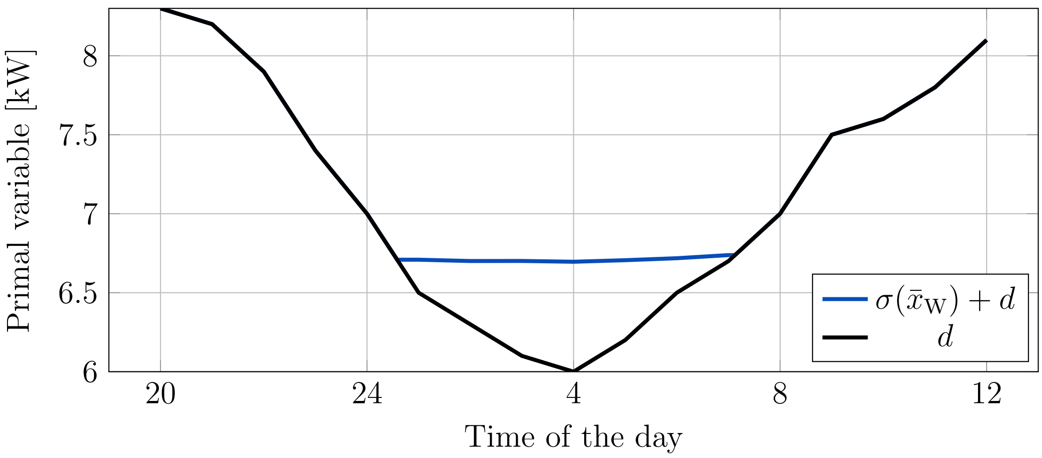


Figure 6.3: The aggregate EV demand $\sigma(\bar{x}_{\mathrm{W}})$ at the Wardrop equilibrium, which is computed by the extragradient algorithm in absence of coupling constraints, exhibits the valley filling property (6.8).

6.1.4 Local coupling constraints.

Instead of the coupling constraint (6.2), we consider here $m = H \cdot n$ local coupling constraints of the form

$$\sum_{j \in \mathcal{N}_h} x_t^j \leq K^h, \forall t \in \{1, \ldots, n\}, \quad \forall h \in \{1, \ldots, H\},$$

where $\mathcal{N}_h \subset \{1, \ldots, M\}$ represents the subset of agents connected to the same transformer h, which cannot provide more than K^h units of power at any time t. Figure 6.4 shows the strategies of three agents, connected to the same transformer, at the variational Nash equilibrium found by Algorithm 4. It is evident that the coupling constraint forces a coordination between agents 1 and 2: agent 2 charges his EV in the first part of the night while agent 1 starts charging in the second part.

6.1.5 Quadratic cost function

Different works in the EV literature [GPC16, KCM11, Gra17] use the quadratic cost of the form (3.31), with $Q \succ 0$ and $C \succ 0$, diagonal. Existence of a Nash and of a Wardrop equilibrium is guaranteed by Proposition 15, while Proposition 16 gives the ε-Nash property. Further, Lemma 4 shows that the resulting operators F_{N} and F_{W} are strongly monotone with monotonicity constant independent from the number of agents M. Theorem 2 ensures then that $\|\bar{x}_{\mathrm{N}} - \bar{x}_{\mathrm{W}}\| \leq L_2/(\alpha\sqrt{M})$, with L_2 that can be shown to be equal to $R \cdot \lambda_{\max}(C)$. A Nash equilibrium can be found using Algorithm 4 (as F_{N} is strongly monotone), while a Wardrop equilibrium can be achieved using both Algorithm 4 (as F_{W} is strongly monotone) and Algorithm 3 (see [GPC16, Theorem 2]).

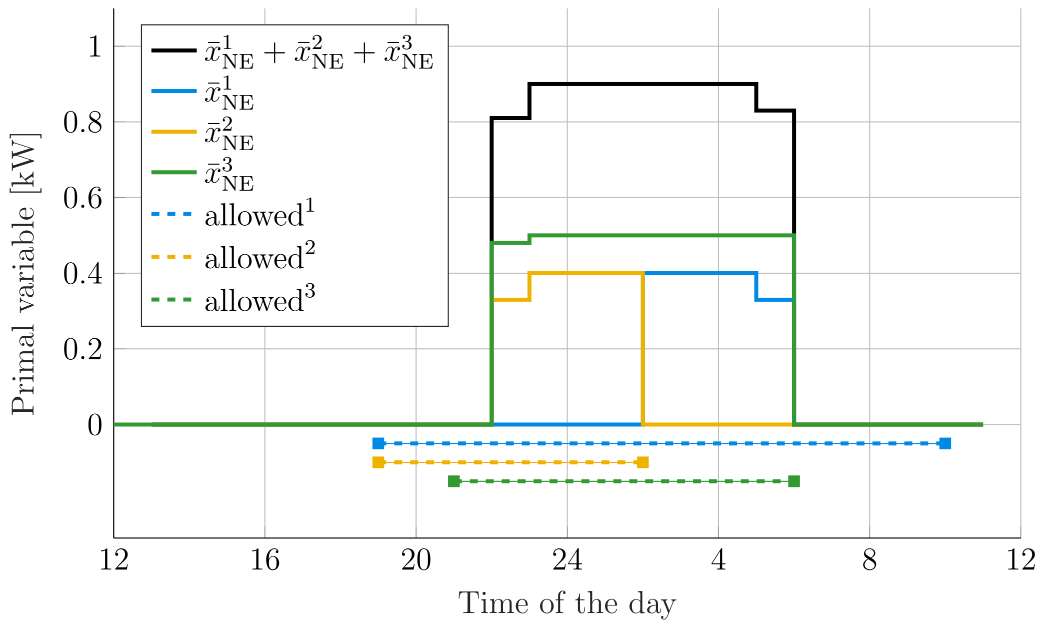


Figure 6.4: Nash strategies $\bar{x}^1_{\mathrm{N}}$, $\bar{x}^2_{\mathrm{N}}$, $\bar{x}^3_{\mathrm{N}}$ and their aggregate $\bar{x}^1_{\mathrm{N}} + \bar{x}^2_{\mathrm{N}} + \bar{x}^3_{\mathrm{N}}$, found with Algorithm 4. The agents differ in θ^i and the local coupling constraint imposes $\bar{x}^1_{\mathrm{N},t} + \bar{x}^2_{\mathrm{N},t} + \bar{x}^3_{\mathrm{N},t} \leq 0.9\mathrm{kW}, \forall t \in \{1, \ldots, n\}$. We also report the interval of allowed charge for each agent i. Outside the allowed interval $\tilde{x}^i_t = 0$, while inside $\tilde{x}^i_t = 1$ for all t, so that the upper bound is inactive, due to the coupling constraint.

Figure 6.5 presents a comparison between Algorithm 3 and 4 for Wardrop in terms of iteration count, where $Q = 0.1I_n$, $C = I_n$, $c^i = d$ for all i. Figure 6.5 (top) represents the number of primal updates required to converge, while Figure 6.5 (bottom) depicts the number of dual updates. For both algorithms the number of iterations does not seem to increase with the number of agents M. Algorithm 3 needs much fewer dual iterations, while Algorithm 4 requires fewer primal iterations, as one would expect given that Algorithm 3 features an inner loop of primal updates for each dual update.

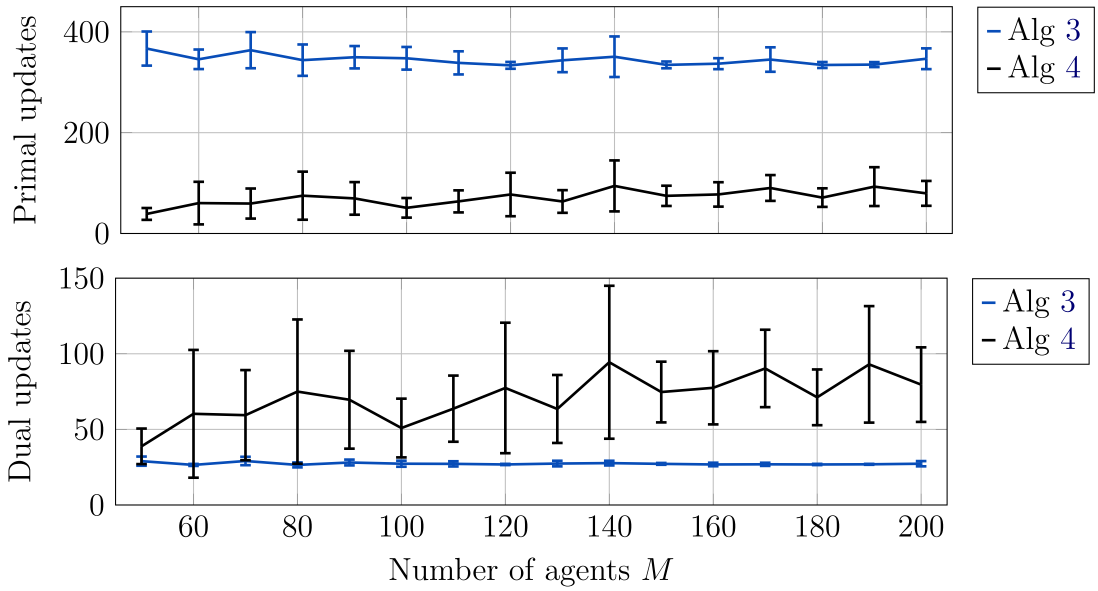


Figure 6.5: Primal (top) and dual (bottom) updates required to converge with precision $\|(x_{(k+1)}, \lambda_{(k+1)}) - (x_{(k)}, \lambda_{(k)})\|_\infty \leq 10^{-4}$; mean and standard deviation for 10 repetitions. As each iteration of Algorithm 4 performs one primal and one dual update, the two black lines (top and bottom) coincide.

6.2 Route choice in a road network

Traffic congestion is a well-recognized issue in densely populated cities, and the corresponding economic costs are significant [AS94]. Traffic engineers have studied the problem for decades from different perspectives, including theoretical [Vic69] and numerical [KHR02] analysis or field experiments [GD08]. Since every driver seeks his own interest (e.g., minimizing the travel time) and is affected by the others' choices via congestion, a classic approach is to model the traffic problem as a noncooperative game [Daf80]. Specializing [FP03, Section 1.4.5], we focus on a stationary model that aims at capturing the basic interactions among the vehicles flow during rush hours. Such model extends the parallel roads network of Example 2 to a generic network, and it differs from [CSM11] in the cost function (6.11), where we introduce a term penalizing the deviation from a preferred route. We assume that the travel time on each road depends only on the traffic on that road, whereas [Daf80] considers also upstream and downstream influence. While most traffic literature focuses solely on the Wardrop equilibrium [CSM11, Daf80], we also study the Nash equilibrium and illustrate the distance between the two. On the other hand, many game theory works on traffic [RT02, CK05] investigate the so-called price of anarchy, or price of stability, whereas we do not focus on the social cost of Nash and Wardrop equilibria. We perform a numerical analysis based on the data set of the city of Oldenburg in Germany [Bri02]. Specifically, we investigate via simulation the effect of road access limitations, expressed as coupling constraints [San75].

Constraints

We consider a strongly-connected directed graph $(\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V} = \{1, \dots, V\}$, representing geographical locations, and directed edge set $\mathcal{E} = \{1, \dots, E\} \subseteq \mathcal{V} \times \mathcal{V}$, representing roads connecting the locations. Each agent $i \in \{1, \dots, M\}$ is a driver who wants to drive from his origin $o^i \in \mathcal{V}$ to his destination $d^i \in \mathcal{V}$.

Let us introduce the vector $x^i \in [0, 1]^E$ to describe the strategy (route choice) of agent i, with $[x^i]_e$ representing the probability that agent i transits on edge e [DPP05]. To guarantee that agent i leaves his origin and reaches his destination with probability 1, the strategy x^i has to satisfy

$$\sum_{e \in \mathrm{in}(v)} [x^i]_e - \sum_{e \in \mathrm{out}(v)} [x^i]_e = \begin{cases} -1 & \text{if} \quad v = o^i \\ 1 & \text{if} \quad v = d^i \\ 0 & \text{otherwise,} \end{cases} \qquad \forall\, v \in \mathcal{V},$$

where $\mathrm{in}(v)$ and $\mathrm{out}(v)$ represent the set of in-edges and the set of out-edges of node v. We denote the graph incidence matrix [Bul18, Chapter 8] by $B \in \mathbb{R}^{V \times E}$, so that $[B]_{ve} = 1$ if edge e points to vertex v, $[B]_{ve} = -1$ if edge e exits vertex v and $[B]_{ve} = 0$

otherwise. The individual constraint set of agent i is then

$$\mathcal{X}^i := \{x \in [0,1]^E : Bx = b^i\}, \tag{6.9}$$

where $b^i \in \mathbb{R}^V$ is such that $[b^i]_v = -1$ if $v = o^i$, $[b^i]_v = 1$ if $v = d^i$ and $[b^i]_v = 0$ otherwise. We introduce the coupling constraint

$$x \in \mathcal{C} := \{x \in \mathbb{R}^{ME} \mid \tfrac{1}{M}\textstyle\sum_{i=1}^{M} x_e^i \le K_e,\ \forall\, e \in \mathcal{E}\}, \tag{6.10}$$

expressing the fact that the number of agents on edge e cannot exceed MK_e. Such constraint can be imposed by authorities [ADPL94] to decrease the congestion in a specific road or neighborhood, with the goal of reducing noise or pollution.

Cost function

We assume that each agent $i \in \{1, \dots, M\}$ wants to minimize his travel time and, at the same time, is not willing to deviate too much from a preferred route $\tilde{x}^i \in \mathcal{X}^i$. We model such trade-off with the following cost function

$$J^i(x^i, \sigma(x)) = \frac{\gamma^i}{2}\|x^i - \tilde{x}^i\|^2 + \sum_{e=1}^{E} t_e(\sigma_e(x_e))x_e^i, \tag{6.11}$$

with $\gamma^i \ge 0$ a weighting factor, $x_e := [x_e^1, \dots, x_e^M]^\top \in \mathbb{R}^M$, $\sigma_e(x_e) = \frac{1}{M}\sum_{i=1}^{M} x_e^i$ and $t_e(\sigma_e(x_e))$ the travel time on edge e. Note that in case agent i travels along a path with probability 1, i.e., $x^i \in \{0,1\}^E$, then $\sum_{e=1}^{E} t_e(\sigma_e(x_e))x_e^i$ represents the expected travel time along that path.

Travel time

This subsection is devoted to the derivation of the analytical expression of the travel time $t_e(\sigma_e(x_e))$. The reader not interested in the technical details of the derivation can jump to the expression of $t_e(\sigma_e(x_e))$ in (6.14), which is illustrated in Figure 6.6. We note that the model under consideration is static because the variable x_e^i denotes the probability that agent i drives through edge e, with no notion of time causality and sequentiality among the edges. Such a model is appropriate to study a road network in a time interval in which the traffic conditions can be considered stationary, as for instance during rush hours. We introduce the quantity $D_e(x_e) = \sum_{i=1}^{M} x_e^i$ to describe the total demand on edge e. We consider a rush-hour interval $[0, h]$ and we assume that the instantaneous demand equals $D_e(x_e)/h$ at any time $t \in [0, h]$ and zero for $t > h$. We assume that edge e can support a maximum flow F_e (agents per unit of time) and features a free-flow travel time $t_{e,\text{free}}$. As we are interested in comparing games with different number of agents, we further assume that the peak hour duration h is independent from the number of agents M and that the edge maximum capacity flow F_e scales linearly with it, i.e. $F_e(M) = f_e \cdot M$, with f_e constant in M. The consideration underpinning this last assumption is that the road infrastructure scales with the number of agents to accommodate the increasing demand, similarly as what assumed in [MCH13] for the energy infrastructure.

If $D_e(x_e)/h \leq F_e$ then every agent has instantaneous access to edge e and no queue accumulates, hence the travel time equals $t_{e,\text{free}}$. We focus in the rest of this paragraph on the case $D_e(x_e)/h > F_e$. An increasing queue forms in the interval $[0, h]$ and, since after the time $t = h$ no more agents arrive, for $t > h$ the queue accumulated until $t = h$ decreases at rate F_e. The number of agents $q_e(t)$ queuing on edge e at time t obeys then the dynamics

$$\dot{q}_e(t) = \begin{cases} \frac{D_e(x_e)}{h} \cdot \mathbf{1}_{[0,h]}(t) - F_e & \text{if } q_e(t) \geq 0 \\ 0 & \text{otherwise,} \end{cases} \qquad q_e(0) = 0, \tag{6.12}$$

where $\mathbf{1}_{[0,h]}$ is the indicator function of $[0, h]$. The solution $q_e(t)$ to (6.12) is hence

$$q_e(t) = \begin{cases} \left(\frac{D_e(x_e)-F_e h}{h}\right) t & \text{if } \ 0 \leq t \leq h, \\ D_e(x_e) - F_e\, t & \text{if } \ h \leq t \leq D_e(x_e)/F_e, \\ 0 & \text{if } \ t \geq D_e(x_e)/F_e. \end{cases} \tag{6.13}$$

As a consequence, the total queuing time at edge e (i.e, the queuing times summed over all agents) is the integral of $q_e(t)$, which equals $D_e(x_e)(D_e(x_e) - F_e h)/(2F_e)$; the (average) queuing time is the total queuing time divided by the total demand, that is, $(D_e(x_e) - F_e h)/(2F_e)$.

As a conclusion, the travel time function for edge e takes the expression

$$t_e^{\text{PWA}}(D_e(x_e)) = \begin{cases} t_{e,\text{free}} & \text{if } \ 0 \leq D_e(x_e) \leq F_e h, \\ t_{e,\text{free}} + \frac{D_e(x_e)-F_e h}{2F_e} & \text{if } \ F_e h \leq D_e(x_e) \leq M. \end{cases}$$

Since $\sigma_e(x_e) = \frac{1}{M}\sum_{i=1}^{M} x_e^i = \frac{1}{M} D_e(x_e)$, we can express the travel time function t_e^{PWA} in terms of $\sigma_e(x_e)$, and with a small abuse of notation write

$$t_e^{\text{PWA}}(\sigma_e(x_e)) = \begin{cases} t_{e,\text{free}} & \text{if } \ 0 \leq \sigma_e(x_e) \leq f_e h, \\ t_{e,\text{free}} + \frac{\sigma_e(x_e)-f_e h}{2f_e} & \text{if } \ f_e h \leq \sigma_e(x_e) \leq 1. \end{cases}$$

The function $t_e^{\text{PWA}}(\sigma_e)$ is reported in Figure 6.6. In words, if the average number of agents on edge e is smaller than a critical threshold, then the travel time equals the free-flow travel time; if instead this number exceeds the critical threshold, then the travel time also accounts for the time spent queuing. Note that t_e^{PWA} is a continuous and piece-wise affine function of σ_e, but it is not continuously differentiable, hence Assumption 1 would not hold. Therefore, we define t_e appearing in (6.11) as the smoothed version of $t_e^{\rm PWA}$

$$t_e(\sigma_e(x_e)) = \begin{cases} t_{e,\rm free} & \text{if } 0 \le \sigma_e(x_e) \le f_e h - \Delta_e, \\ \frac{1}{2} a \sigma_e(x_e)^2 + b\sigma_e(x_e) + c & \text{if } f_e h - \Delta_e \le \sigma_e(x_e) \le f_e h + \Delta_e, \\ t_{e,\rm free} + \frac{\sigma_e(x_e) - f_e h}{2 f_e} & \text{if } f_e h + \Delta_e \le \sigma_e(x_e) \le 1. \end{cases} \tag{6.14}$$

where the values of Δ_e, a, b, c are such that t_e is continuously differentiable[2], as illustrated in Figure 6.6. We note that the function $t_e(\sigma_e(x_e))$ is used within a stationary

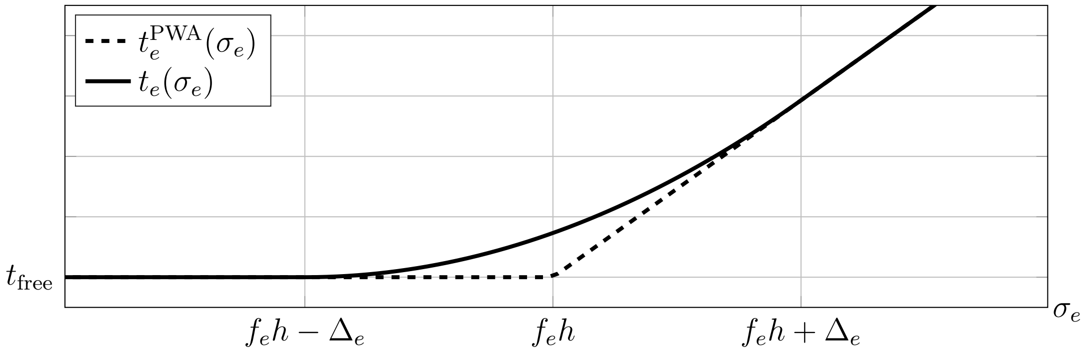


Figure 6.6: Piece-wise affine travel time $t_e^{\rm PWA}(\sigma_e)$ and its smooth approximation $t_e(\sigma_e)$.

traffic model but includes the average queuing time which is based on the dynamic function (6.13). We have performed a more thorough analysis of a dynamic traffic model in [BPG17], but we remark that the dynamic traffic equilibrium is in general known to be hard to compute [LS02].

Finally, we remark that a travel time with similar monotonicity properties can be derived from the piece-wise affine fundamental diagram of traffic [LZ11, Figure 7], but $t_e(\sigma_e(x_e))$ would present a vertical asymptote which is absent here.

6.2.1 Theoretical guarantees

We define the route-choice game $\mathcal{G}_M^{\rm RC}$ as in (3.3), with (6.9) defining $\mathcal{X}^i$, (6.10) specifying $\mathcal{C}$, and (6.11), (6.14) constituting $J^i(x^i, \sigma(x))$. In the following we specialize the main results of Chapters 3 and 4 to the route choice game.

Corollary 4. Consider the sequence of games $(\mathcal{G}_M^{\rm RC})_{M=1}^\infty$. Assume that for each game $\mathcal{G}_M^{\rm RC}$ the set $\mathcal{Q} = \mathcal{C} \cap \mathcal{X}$ is non-empty, that $h > 0$ and $t_{e,\rm free}, f_e > 0$ for each $e \in \mathcal{E}$. Moreover, assume that there exists $\hat{\gamma} > 0$ such that $\gamma^i \ge \hat{\gamma}$ for all $i \in \{1, \dots, M\}$, for all M. Then:

[2]The values are $\Delta_e = 0.5(\sqrt{(f_e h)^2 + 4 f_e h} - f_e h)$, $a = 1/(4 f_e \Delta_e)$, $b = 1/(4 f_e) - h/(4\Delta_e)$, $c = t_{e,\rm free} + (f_e h)^2/(8 f_e \Delta_e) - h/4 - (\Delta_e)/(8 f_e)$.

1. The operator F_{W} is strongly monotone, hence each game $\mathcal{G}_M^{\mathrm{RC}}$ admits a unique variational Wardrop equilibrium. For every M satisfying

$$M > \max_{e\in\mathcal{E}} \frac{1}{32 f_e \Delta_e \hat{\gamma}} \tag{6.15}$$

the operator F_{N} is strongly monotone, hence each game $\mathcal{G}_M^{\mathrm{RC}}$ admits a unique variational Nash equilibrium. Every Wardrop equilibrium is an ε-Nash equilibrium with $\varepsilon = \frac{E}{M f_{\min}}$, where $f_{\min} = \min_{e\in\mathcal{E}} f_e$.

2. For any variational Nash equilibrium $\bar{x}_{\mathrm{N}}$ of $\mathcal{G}_M^{\mathrm{RC}}$, the unique variational Wardrop equilibrium $\bar{x}_{\mathrm{W}}$ of $\mathcal{G}_M^{\mathrm{RC}}$ satisfies

$$\|\bar{x}_{\mathrm{N}} - \bar{x}_{\mathrm{W}}\| \leq \frac{\sqrt{E}}{2 f_{\min} \hat{\gamma} \sqrt{M}}.$$

3. For any M, Algorithm 4 with operator F_{W} converges to a variational Wardrop equilibrium of $\mathcal{G}_M^{\mathrm{RC}}$. For M satisfying (6.15), Algorithm 4 with operator F_{N} converges to a variational Nash equilibrium of $\mathcal{G}_M^{\mathrm{RC}}$. □

Proof. 1) Assumption 1 and the consequent existence of a variational Nash and of a variational Wardrop equilibrium for any M can be shown as in Corollary 3. The operator F_{W} for the cost (6.11) reads

$$F_{\mathrm{W}}(x) = [\gamma^i(x^i - \hat{x}^i) + t(\sigma(x))]_{i=1}^M, \tag{6.16}$$

where $t(\sigma(x)) \coloneqq [t_e(\sigma_e(x_e))]_{e=1}^E$. Since $t_e(\sigma_e(x_e))$ in (6.14) is a monotone function of $\sigma_e(x_e)$, it is straightforward to show that the operator $t(\sigma(x))$ is monotone in x. Then F_{W} is strongly monotone with constant $\hat{\gamma}$ because it is the sum of a monotone and a strongly monotone operator with constant $\hat{\gamma}$ by Lemma 1. As a consequence, each $\mathcal{G}_M^{\mathrm{RC}}$ admits a unique variational Wardrop equilibrium.

We now show that under (6.15) F_{N} is strongly monotone. By (3.20b), its expression is

$$F_{\mathrm{N}} = F_{\mathrm{W}} + \frac{1}{M}[[t'_e(\sigma_e(x_e))x_e^i]_{e=1}^E]_{i=1}^M. \tag{6.17}$$

To prove strong monotonicity of F_{N} it suffices to show that the term $[[t'_e(\sigma_e(x_e))x_e^i]_{e=1}^E]_{i=1}^M$ is monotone, as F_{W} is strongly monotone. However, to this end we cannot directly use Theorem 1, because t_e is not strictly increasing in the interval $[0, f_e h - \Delta_e]$ and because t is not twice continuously differentiable[3]. For this reason, we conduct a slightly different proof. Under the assumptions of Proposition 8, monotonicity of $[[t'_e(\sigma_e(x_e))x_e^i]_{e=1}^E]_{i=1}^M$ is guaranteed if

$$\nabla_x \left([[t'_e(\sigma_e(x_e))x_e^i]_{e=1}^E]_{i=1}^M\right) \succeq 0, \quad \forall x \in [0,1]^{ME}, \tag{6.18}$$

[3]To see it, observe that t_e is piece-wise quadratic, continuous with continuous derivatives, but $t''_e(\sigma_e) = 0$ for $\sigma_e \in [0, f_e h - \Delta_e]$ or $\sigma_e > f_e h + \Delta_e$, while $t''_e(\sigma_e) = 1/(8 f_e \Delta_e) > 0$ for $\sigma_e \in [f_e h - \Delta_e, f_e h - \Delta_e]$.

because $\mathcal{Q} = \mathcal{X} \cap \mathcal{C} \subset [0,1]^{ME}$. Condition (6.18) is equivalent to

$$\nabla_{x_e}\left([t'_e(\sigma_e(x_e))x^i_e]_{i=1}^M\right) \succeq 0, \quad \forall x_e \in [0,1]^M,\ \forall e \in \mathcal{E}. \tag{6.19}$$

The issue here is that Proposition 8 cannot be applied, as it requires F_{N} to be continuously differentiable and this is not the case because t is not twice continuously differentiable, as observed above. Nonetheless, we can use [DTL96, Proposition 2.1], which extends the classical result of Proposition 8 to operators which are not necessarily continuously differentiable, but at least admit a generalized Jacobian (which is the multidimensional version of the generalized-gradient, or sub-gradient, see [Cla90, Definition 2.6.1]). In our setup, it is crucial to observe that $t'_e(\sigma_e)$ is continuously differentiable in each of the open intervals $(0, f_e h - \Delta_e)$, $(f_e h - \Delta_e, f_e h + \Delta_e)$, and $(f_e h + \Delta_e, 1)$. Then each matrix belonging to the sub-Jacobian at the points $\sigma_e = f_e h - \Delta_e$ and $\sigma_e = f_e h + \Delta_e$ is positive semi-definite if the Jacobian is positive semi-definite in the three intervals separately (and the same holds for the points $\sigma_e = 0$ and $\sigma_e = 1$). Mathematically this means that instead of condition (6.19) it suffices to check, for all $e \in \mathcal{E}$, the following three conditions separately.

$$\nabla_{x_e}\left([t'_e(\sigma_e(x_e))x^i_e]_{i=1}^M\right) \succeq 0, \forall x_e \in [0,1]^M \text{ s.t. } \sigma_e(x_e) \in (0, f_e h - \Delta_e) \tag{6.20a}$$

$$\nabla_{x_e}\left([t'_e(\sigma_e(x_e))x^i_e]_{i=1}^M\right) \succeq 0, \forall x_e \in [0,1]^M \text{ s.t. } \sigma_e(x_e) \in (f_e h - \Delta_e, f_e h + \Delta_e) \tag{6.20b}$$

$$\nabla_{x_e}\left([t'_e(\sigma_e(x_e))x^i_e]_{i=1}^M\right) \succeq 0, \forall x_e \in [0,1]^M \text{ s.t. } \sigma_e(x_e) \in (f_e h + \Delta_e, 1). \tag{6.20c}$$

Conditions (6.20a) and (6.20c) are trivially satisfied, because in those two intervals $t'_e(\sigma_e)$ is constant, hence $\nabla_{x_e}\left([t'_e(\sigma_e(x_e))x^i_e]_{i=1}^M\right)$ is a multiple of the identity.

To verify Condition (6.20b) we can now legitimately invoke Theorem 1, because indeed $[t'_e(\sigma_e(x_e))x^i_e]_{i=1}^M$ is continuously differentiable in $(f_e h - \Delta_e, f_e h + \Delta_e)$. Specifically, we can use a slight modification of the sufficient condition (3.23), where it is enough to show that the left-hand side is greater than $-\hat{\gamma}M$, because of (6.17) and the fact that F_{W} features strong monotonicity constant $\hat{\gamma}$ by (6.16). This translates into the sufficient condition

$$\min_{\substack{e\in\mathcal{E} \\ \sigma_e \in (f_e h - \Delta_e, f_e h + \Delta_e)}} \left(t'_e(\sigma_e) - \frac{t''_e(\sigma_e)}{8}\right) > -\hat{\gamma}M,$$

which, since $t'(\sigma_e) \geq 0$ and $t''_e(\sigma_e) = 1/(4 f_e \Delta_e)$, is implied by

$$\max_{e\in\mathcal{E}} \frac{1}{32\, f_e \Delta_e} < \hat{\gamma}M,$$

which is equivalent to (6.15). We can conclude that under condition (6.15) F_{N} is strongly monotone and thus $\mathcal{G}_M^{\mathrm{RC}}$ admits a unique variational Nash equilibrium.

Finally, we verify Assumption 2 in order to use Proposition 16 that guarantees the ε-Nash property. We have $\mathcal{X}^0 = [0,1]^E$ and t is continuously differentiable hence Lipschitz in $\mathcal{X}^0$, with Lipschitz constant $L_p = 1/(2f_{\min})$ (see (6.14) or Figure 6.6). Moreover, $R := \max_{y\in\mathcal{X}^0}\{\|y\|\} = \|\mathbb{1}_E\| = \sqrt{\sum_{e\in\mathcal{E}} 1} = \sqrt{E}$. Then (3.35) establishes that L_2 in Proposition 16 equals RL_p, which in our case reads $L_2 = \sqrt{E}/(2f_{\min})$. Hence we can conclude that the quantity $\varepsilon = 2RL_2/M$ of Proposition 16 becomes here $\varepsilon = E/(Mf_{\min})$, thus concluding the proof of the first statement.

2) Since all the assumptions of Theorem 2 have just been verified, it is a direct consequence of its second statement. Specifically, by substituting the expressions $L_2 = \sqrt{E}/(2f_{\min})$ and $\alpha_M = \hat{\gamma}$ derived above, the upper bound $L_2/(\alpha_M\sqrt{M})$ in (3.37) reads here $\sqrt{E}/(2f_{\min}\hat{\gamma}\sqrt{M})$.

3) As Assumption 3 holds trivially, all the assumptions of Theorem 4 have just been verified and its statement concludes the proof. □

6.2.2 Numerical analysis

For the numerical analysis we use the data set of the city of Oldenburg [Bri02], whose graph features 175 nodes and 213 undirected edges[4] and is reported in Figure 6.7. For each agent i the origin o^i and the destination d^i are chosen uniformly at random. Regarding the cost (6.11), $t_{e,\text{free}}$ is computed as the ratio between the edge length, which is provided in the data set, and the free-flow speed. Based on the road topology, we divide the roads into main roads, where the free-flow speed is 50 km/h, and secondary roads, where the free-flow speed is 30 km/h. Moreover, we assume a peak hour duration h of 2 hours, and for all $e \in \mathcal{E}$, we set $f_e = 4 \cdot 10^{-3}$ agents per second, which corresponds to 1 vehicle every 4 seconds for $M = 60$ agents. Finally, the parameter γ^i is picked uniformly at random in $[0.5, 3.5]$ and $\tilde{x}^i$ is such that $\tilde{x}^i_e = 1$ if e belongs to the shortest path from o^i to d^i, while $\tilde{x}^i_e = 0$ otherwise. The shortest path is computed based on $\{t_{e,\text{free}}\}_{e=1}^E$. Note that with the above values the bound (6.15) becomes $M > 16.14$, which is satisfied also for a small number of agents.

We compute the Wardrop equilibrium with Algorithm 4 relatively to $M = 60$ drivers without coupling constraint, i.e. with $K_e = 1$ for all $e \in \mathcal{E}$. We report in Figure 6.7 the corresponding queuing time $t_e(\sigma_e(x_e)) - t_{e,\text{free}}$ as by (6.14).

We illustrate in Figure 6.8 the change in the queuing time of an entire neighborhood when introducing a coupling constraint that upper bounds the total number of agents on a single edge, relatively to a Wardrop equilibrium with $M = 60$. At the Wardrop equilibrium, the dual variable λ_e corresponding to the coupling constraint equals 21.2 units of time. Thanks to Proposition 17, this can be interpreted as a tolling price that a vehicle pays to use the road subject to the coupling constraint.

[4]The graph in the original data set features 6105 vertexes and 7035 undirected edges. For reasons of computational tractability, we reduce it by excluding all the nodes that are outside the rectangle $[3619, 4081] \times [3542, 4158]$ and all the edges that do not connect two nodes in the rectangle. The resulting graph is strongly connected.

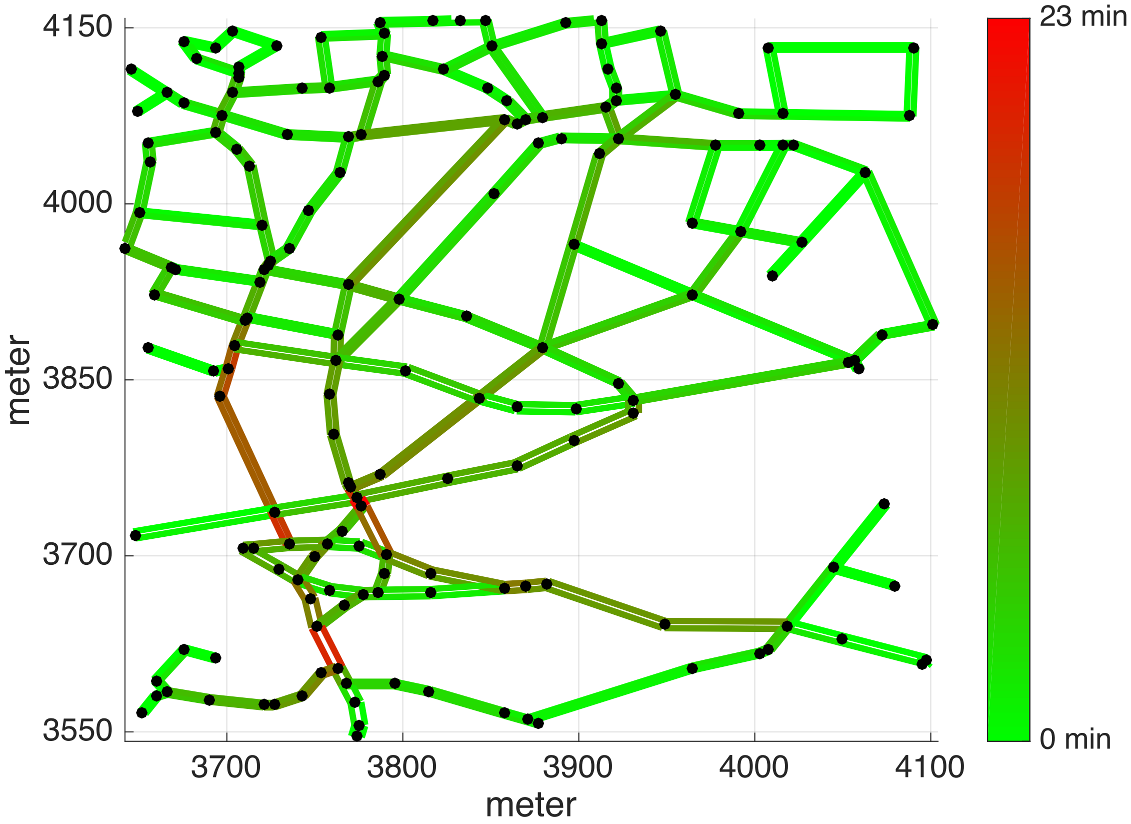


Figure 6.7: The queuing time $t_e(\sigma_e(x_e)) - t_{e,\text{free}}$ reported in green-red color scale. Note that this pattern changes if one modifies the origin-destination pairs. Each edge presents a specific travel direction and the one next to it is used for the opposite travel direction. The convention is that vehicles drive on the right side.

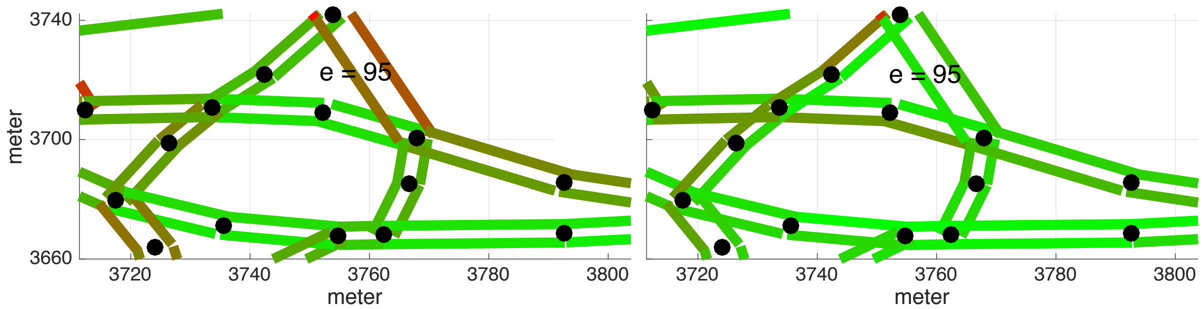


Figure 6.8: On the left, the queuing time in a neighborhood without any coupling constraints; 10% of the agents transits on edge 95, and the queuing time is 7.28 minutes. On the right, the queuing time in presence of a coupling constraint allowing at most 3% of the entire agents on edge 95; the queuing time is reduced to 1.42 minutes, but it visibly increases on the edges of the alternative route.

Finally, we illustrate the second statement of Corollary 4 by reporting in Figure 6.9 the distance between the unique variational Wardrop equilibrium and the variational Nash equilibrium found by Algorithm 4. The ε-Nash property of the Wardrop equilibrium in Proposition 16 can also be illustrated with a similar plot, which is omitted here.

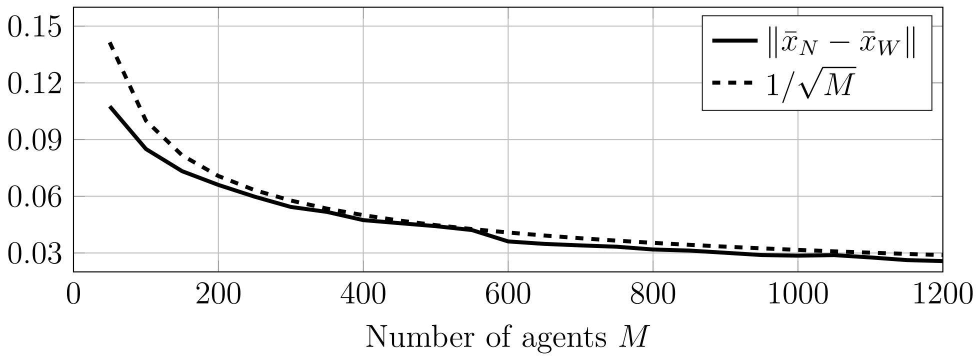


Figure 6.9: Distance between variational Nash and Wardrop equilibria. As in Fig. 6.2, the function $1/\sqrt{M}$ illustrates the trend of the bound derived in Corollary 4 and not the specific constant, which is conservative and not shown here.

6.3 Cournot game with transportation costs

To illustrate the theoretical findings of Chapter 5, we study in detail a Cournot game with transportation costs, as introduced[5] in [FP03, Section 1.4.3]. Such setup extends the Cournot game [Jen10, Section 2] and the multi-market Cournot game [YP17, Section 7.1] because it considers transportation costs. Our novel contributions consist in introducing coupling constraints, providing sufficient conditions for strong monotonicity based on Theorem 1, and focusing on distributed convergence.

Consider a single-commodity[6] Cournot game with M firms and V markets, which correspond to V physical locations. Firm $i \in \{1, \dots, M\}$ chooses to sell $y_v^i \in \mathbb{R}_{\geq 0}$ amount of commodity at each market $v \in \{1, \dots, V\}$. Each firm i produces its commodity at a given fixed location $\ell_i \in \{1, \dots, V\}$ and then ships its commodity to the different markets over a transportation network, where the V nodes represent market locations and a directed edge connecting two nodes represents a road connecting two markets. We characterize the network by its incidence matrix $B \in \{0, 1, -1\}^{V \times E}$, where E is the number of edges and $B_{v,e} = -1$ if edge e leaves node v, $B_{v,e} = 1$ if edge e enters node v and $B_{v,e} = 0$ otherwise [Bul18, Chapter 8]. Denote by $r^i \in \mathbb{R}_{\geq 0}$ the total amount of commodity produced and sold by firm i (i.e., $r^i = \sum_{v=1}^{V} y_v^i$) and by $t_e^i \in \mathbb{R}_{\geq 0}$ the amount of commodity transported by firm i over edge e, with $t^i = [t_e^i]_{e=1}^{E}$. Define the strategy vector of firm i as $x^i := [t^i; r^i] \in \mathbb{R}_{\geq 0}^{E+1}$, which uniquely determines $y^i := [y_v^i]_{v=1}^{V}$, due to the balance equation

$$y^i = Bt^i + \mathbf{e}_{\ell_i} r^i = H^i x^i, \tag{6.21}$$

with $H^i := [B, \mathbf{e}_{\ell_i}] \in \mathbb{R}^{V \times (E+1)}$ and $\mathbf{e}_j$ the j^{th} canonical vector. The balance equation (6.21) states that the commodity y_v^i sold at node $v \neq \ell_i$ equals the commodity shipped to v minus the commodity that from v is shipped further in the network. The commodity sold at ℓ_i equals the difference between the total production r^i and the total amount shipped from the production node ℓ_i.

Cost function

We assume that at each market the commodity is sold at a price that depends on the total commodity sold by the M firms. We allow for inter-market effects and use the inverse demand function[7] $p : \mathbb{R}_{\geq 0}^{V} \to \mathbb{R}_{\geq 0}^{V}$ that maps the normalized vector $\sigma(x) = \frac{1}{M} \sum_{j=1}^{M} y^j = \frac{1}{M} \sum_{j=1}^{M} H^j x^j$ to the vector of prices of each market $p(\sigma(x)) := [p_v(\sigma(x))]_{v=1}^{V}$. We stress that in this section $\sigma(x)$ represents $\frac{1}{M} \sum_{j=1}^{M} H^j x^j$ rather than $\frac{1}{M} \sum_{j=1}^{M} x^j$. Then the

[5] We developed the model from scratch to later find out that it is identical to [FP03, Section 1.4.3].

[6] The analysis applies also to a multi-commodity game as in [YP17, Section 7.1], but we consider a single-commodity game for ease of exposition and we rather focus on the transportation costs.

[7] The inverse demand function determines the price for which demand equals supply at market v. This is why we can assume that all the supply is sold.

revenue of firm i is $p(\sigma(x))^\top y^i$. Moreover, for firm i transporting t_e^i commodity over an edge comes with a cost equal to

$$c_e^i(t_e^i) := \beta_e^i t_e^i - \gamma_e^i(t_e^i), \tag{6.22}$$

where, for all i, γ_e^i is a strongly concave, increasing function with maximum derivative smaller than β_e^i. The transportation cost in (6.22) can be thought of as the sum of two terms: the first is a cost proportional to the amount shipped, the second term is a discount that increases as the amount of shipped commodity increases.

The production cost function of firm i has a similar form

$$a^i(r^i) := \beta_a^i r^i - \gamma_a^i(r^i), \tag{6.23}$$

where γ_a^i is a strongly concave, increasing function with maximum derivative smaller than β_a^i. Note that the functions (6.22) and (6.23) are strongly convex, as in [FP03, Section 1.4.3].

To sum up, the cost function of firm i is

$$J^i(x^i, \sigma(x)) := \underbrace{a^i(r^i)}_{\text{production cost}} + \underbrace{\textstyle\sum_{e=1}^{E} c_e^i(t_e^i)}_{\text{transportation cost}} - \underbrace{p(\sigma(x))^\top y^i}_{\text{revenue}}. \tag{6.24}$$

Constraints

The strategy of firm i must satisfy the individual constraints

$$\mathcal{X}^i := \{x^i \in \mathbb{R}_{\geq 0}^{E+1} | x^i \leq \bar{r}^i \cdot \mathbb{1}_{E+1}, y^i = H^i x^i \geq \mathbb{0}_V\}, \tag{6.25}$$

where $\bar{r}^i$ is the production capacity of firm i. Note that (6.25) implies $t_e^i \leq \bar{r}^i$ for each e, which is needed to guarantee boundedness of $\mathcal{X}^i$ and it can be imposed without loss of generality if $r^i \leq \bar{r}^i$, as the transportation costs c_e^i are increasing.

Moreover, we assume that each market v is composed by retailers whose storage capacity imposes an upper bound $K_v > 0$ on the total commodity that can be sold at market v, thus giving rise to the coupling constraints $\sigma(x) \leq K := [K_v]_{v=1}^V$.

Communication network

We assume that the firms can communicate with each other according to a sparse communication network, described by the adjacency matrix T, which we assume satisfies Assumption 6. This network can model spatial proximity of firms, or the fact that they may want to share their strategies only with firms they trust.

6.3.1 Theoretical guarantees

We focus hereafter on a Nash equilibrium as in Chapter 5; we denote with F the Nash operator, which is instead denoted with F_{N} in Chapters 3 and 4. We define the Cournot game $\mathcal{G}^{\mathrm{CO}}$ as in (3.3), with $\mathcal{X}^i$, $\mathcal{C}$ and $J^i(x^i,\sigma(x))$ as introduced above. The only difference with (3.3) is that the aggregate $\sigma(x)$ depends on $y^i = H^i x^i$ instead of x^i directly. We next show that the statements of Chapter 5 can be easily extended to cover such case[8].

Extension

Set $H_{\mathrm{blkd}} \coloneqq \mathrm{blkdiag}(H^1,\dots,H^M) \in \mathbb{R}^{MV\times M(E+1)}$. The quantities in (5.4) relative to $\mathcal{G}$ generalize to

$$
\begin{aligned}
F(x) &\coloneqq [\nabla_{x^i} J^i(x^i,\sigma(x))]_{i=1}^M,\\
&= [\nabla_{z_1} J^i(x^i,\sigma(x)) + \tfrac{1}{M} H_i^\top \nabla_{z_2} J^i(x^i,\sigma(x))]_{i=1}^M,\\
\mathcal{Q} &\coloneqq \{x\in\mathcal{X}^1\times\cdots\times\mathcal{X}^M | Ax\le b\}, && (6.26a)\\
A &\coloneqq \left(\frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top \otimes \hat{A}\right) H_{\mathrm{blkd}}, && (6.26b)\\
b &\coloneqq \mathbb{1}_M \otimes \hat{b}. && (6.26c)
\end{aligned}
$$

The quantities in (5.3) relative to $\mathcal{G}_\nu$ generalize to

$$
\begin{aligned}
F_\nu(x) &\coloneqq [\nabla_{x^i} J^i(x^i,\sigma_\nu^i(x))]_{i=1}^M, && (6.27a)\\
&= [\nabla_{z_1} J^i(x^i,\sigma_\nu^i(x)) + [T^\nu]_{ii} H_i^\top \nabla_{z_2} J^i(x^i,\sigma_\nu^i(x))]_{i=1}^M,\\
\mathcal{Q}_\nu &\coloneqq \{x\in\mathcal{X}^1\times\cdots\times\mathcal{X}^M | A_\nu x\le b\}, && (6.27b)\\
A_\nu &\coloneqq (T^\nu\otimes\hat{A})H_{\mathrm{blkd}}. && (6.27c)
\end{aligned}
$$

As the coupling constraint $Ax\le b$ cannot be expressed on the average $\sigma(x)$ because of the presence of H_{blkd} in (6.26b), we replace Assumptions 5 and 7 with the following one.

Assumption 7'. The matrix A in (6.26b) and the vector b in (6.26c) are such that the following implication holds.

$$
\{A^\top s = \mathbb{0}_{Mn},\quad b^\top s\le 0,\quad s\ge \mathbb{0}_m\} \quad\Rightarrow\quad s=\mathbb{0}_m. \qquad \square
$$

The proofs of Theorems 5 and 6, which rely on Assumptions 5 and 7, can be conducted in the same way using Assumption 7'.

[8]Defining a new game with strategies $\tilde{x}^i = H^i x^i$, so that the aggregate depends only on the $\tilde{x}^i$, is not possible in general as the cost $J^i(x^i,\sigma(x))$ cannot be expressed as a function of the variables $(\tilde{x}^i)_{i=1}^M$ unless H^i is full column rank for all i. This is not the case in the Cournot game under consideration.

Verify the assumptions

To use Theorem 5 in the numerical analysis of the next subsection, we need to verify its Assumptions 1, 6, 7', 8. Assumption 6 holds by problem statement. Assumption 7' is satisfied, because $\{(\mathbb{1}_M \otimes K)^\top s \leq 0, s \geq \mathbb{0}_V\} \Rightarrow s = \mathbb{0}_V$, as $K > \mathbb{0}_V$. To guarantee Assumption 8 we make the following assumption, whose sufficiency is proven in Lemma 12.

Assumption 10 (Cournot-game regularity conditions)**.** The cost $J^i(z_1, z_2)$ is twice continuously differentiable in $[z_1; z_2]$ for all i, and the inverse demand function p satisfies one of the following conditions.
1) p is affine, i.e., $p(\sigma(x)) = -D\sigma(x) + d$, for some $D \in \mathbb{R}^{V\times V}$, $d \in \mathbb{R}^V$ and $D \succeq 0$.
2) p_v depends only on the commodity sold at v, i.e., $p(\sigma(x)) =: [p_v(\sigma_v(x))]_{v=1}^V$. For each v, p_v is twice continuously differentiable, strictly decreasing and satisfies

$$\min_{\substack{v\in\{1,\dots,V\}\\ z\in[0,\tilde{r}]}} \left(-p_v'(z) + \frac{\tilde{r}p_v''(z)}{8}\right) > 0, \quad \tilde{r} := \max_{i\in\{1,\dots,M\}} \bar{r}^i. \tag{6.28}$$

Lemma 12. Under Assumption 10 the Cournot game $\mathcal{G}^{\mathrm{CO}}$ satisfies Assumption 8.

Proof. The set $\mathcal{X}$ is bounded as $\mathcal{X} \subseteq [0, \tilde{r}]^{M(E+1)}$ and J^i is twice continuously differentiable by assumption, so we are left with proving strong monotonicity of F_{N}. We start by expressing the operator F as

$$F(x) = [\nabla_{x^i}\left(a^i(r^i) + \sum_{e=1}^{E} c_e^i(t_e^i)\right)]_{i=1}^M + P(x),$$

where $P(x) := -[\nabla_{x^i}(p(\sigma(x))^\top y^i)]_{i=1}^M$. Since for each i the functions a^i and c_e^i are strongly convex and continuously differentiable, by Table 2.1 and Proposition 8 there exists $\alpha > 0$ such that

$$\nabla_x([\nabla_{x^i}\left(a^i(r^i) + \sum_{e=1}^{E} c_e^i(t_e^i)\right)]_{i=1}^M) \succ \alpha I_{M(E+1)}, \ \forall\, x \in \mathcal{X}.$$

We now prove that $\nabla_x P(x) \succeq 0$ under either of the two conditions stated.
1) We have

$$\begin{aligned} M \cdot P(x) &:= \Big[\sum_j H_i^\top D H_j x^j + H_i^\top D^\top H_i x^i - M H_i^\top d\Big]_{i=1}^M \\ &= [H^\top D H + H_{\mathrm{blkd}}^\top D^\top H_{\mathrm{blkd}}]x - M[H_i^\top d]_{i=1}^M, \end{aligned}$$

with $H := ([H_i^\top]_{i=1}^M)^\top$ and $H_{\mathrm{blkd}} = \mathrm{blkdiag}(H_1, \dots, H_M)$. Moreover, since $D \succeq 0$, then

$$\nabla_x P(x) = \frac{1}{2M}(H^\top (D + D^\top) H + H_{\mathrm{blkd}}^\top (D + D^\top) H_{\mathrm{blkd}}) \succeq 0.$$

2) Let $\tilde{P}(y) := -[\nabla_{y^i}(p(\frac{1}{M}\sum_{j=1}^{M} y^j)^\top y^i)]_{i=1}^M$. By Theorem 1, under (6.28) $\tilde{P}(y)$ is strongly monotone, i.e., there exists $\alpha' > 0$ such that $\nabla_y \tilde{P}(y) \succ \alpha' I_{ME}$ by Proposition 8. Note that condition (6.28) exhibits a minus sign compared to condition (3.23), as p appears with a minus sign in (6.24) and with a plus sign in (3.22). Moreover, from $p(\sigma(x))^\top y^i = p(\frac{1}{M}\sum_{j=1}^{M} H^j x^j)^\top H^i x^i$ one immediately gets that $P(x) = H_{\text{blkd}}^\top \tilde{P}(H_{\text{blkd}} x)$. It follows that for any x and corresponding $y = H_{\text{blkd}} x$,

$$\nabla_x P(x) = (H_{\text{blkd}}^\top \nabla_y \tilde{P}(y) H_{\text{blkd}})_{|y=H_{\text{blkd}}x} \succeq 0. \tag{6.29}$$

We have proven that $\nabla_x F(x) \succ \alpha I_{M(E+1)}$ for all x. Consequently, F is strongly monotone by Proposition 8 and Assumption 8 is satisfied. □

Remark 5. If the function p is as in Assumption 10.1 and $D = D^\top$, then $\mathcal{G}^{\text{CO}}$ is a potential game [MS96]. In other words, there exists a function $f : \mathcal{Q} \to \mathbb{R}$ such that $\nabla_x f(x) = F(x)$ and VI($\mathcal{Q}$,F) is equivalent to $\operatorname{argmin}_{x\in\mathcal{Q}} f(x)$, as described in Figure 2.1. Then a Nash equilibrium can be found by solving the optimization program $\operatorname{argmin}_{x\in\mathcal{Q}} f(x)$. We also point out that condition (6.28) is satisfied if for each market v the function p_v is convex and strictly decreasing. □

Regarding Assumption 1, Assumption 10 implies that J^i is continuously differentiable in its arguments and that $\nabla_x[\nabla_{x^i} J^i(x^i, \sigma(x))]_{i=1}^M \succ \alpha I_{M(E+1)}$ by Proposition 8, which in turn implies $\nabla_{x^i}\nabla_{x^i} J^i(x^i, \sigma(x)) \succ \alpha I_{E+1}$, which implies convexity of J^i in x^i for all fixed x^{-i}. The sets $\mathcal{X}^i$ are trivially convex, closed and have non-empty interior.

Finally, the next lemma shows that for the network, number of communications ν and price functions used in the numerical analysis of the next subsection, F_ν is strongly monotone, as required by Theorem 7.

Lemma 13. Under Assumption 10.1, if T is the adjacency matrix of an undirected network, so that $T = T^\top$, then the operator F_ν is strongly monotone for any ν even.

Proof. The expression of $F_\nu(x)$ is very similar to $F(x)$ in Lemma 12

$$F_\nu(x) = [\nabla_{x^i}\left(a^i(r^i) + \sum_{e=1}^{E} c_e^i(t_e^i)\right)]_{i=1}^M + P_\nu(x),$$

where $P_\nu(x) := -[\nabla_{x^i}(p(\sigma_\nu^i(x))^\top y^i)]_{i=1}^M$. Since F_ν is continuously differentiable, we can prove its strong monotonicity by showing that there exists $\alpha > 0$ such that $\nabla_x F_\nu(x) \succ \alpha I_{M(E+1)}$, thanks to Proposition 8. As in Lemma 12, there exists $\alpha > 0$ such that $\nabla_x[\nabla_{x^i}\left(a^i(r^i) + \sum_{e=1}^{E} c_e^i(t_e^i)\right)]_{i=1}^M \succ \alpha I_{M(E+1)}$ for all x, hence the proof is concluded upon showing that for all ν even, $\nabla_x P_\nu(x) \succeq 0$. If we denote $\tilde{P}_\nu(y) := -[\nabla_{y^i}(p(\sigma_\nu^i(y))^\top y^i)]_{i=1}^M$ then simple algebraic computations show that

$$\nabla_y \tilde{P}_\nu(y) = \text{blkdiag}([T^\nu]_{11}D, \ldots, [T^\nu]_{MM}D) + T^\nu \otimes D. \tag{6.30}$$

Note that $[T^\nu]_{ii} \geq 0$ for all i and for all ν, hence the first summand in (6.30) is positive semidefinite. Note that, as the matrix T is symmetric, $T^\nu \succeq 0$ for all ν even and thus $T^\nu \otimes D \succeq 0$ since it is the Kronecker product of two symmetric positive semidefinite matrices. Overall, we have $\nabla_y \tilde{P}_\nu(y) \succeq 0$ for all ν even. Finally, as in (6.29), we have

$$\nabla_x P_\nu(x) = (H_{\text{blkd}}^\top \nabla_y \tilde{P}_\nu(y) H_{\text{blkd}})_{|y=H_{\text{blkd}}x} \succeq 0. \qquad \square$$

6.3.2 Numerical analysis

We consider two simulation setups. The first is a small example to develop intuition about the problem, the second is used to illustrate the applicability to a more complex scenario.

Small network

We consider a simple chain transportation network with $V = 5$ markets, $E = 4$ roads and $M = 3$ firms. As illustrated in Figure 6.10 we assume that the firms $\{1, 2, 3\}$ are located at markets $\{1, 3, 5\}$, respectively, and are otherwise identical, with $\bar{r}^i = 5$ for all i.

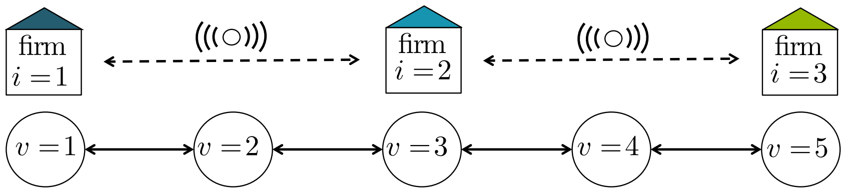


Figure 6.10: Small network with the 3 firms located in markets 1, 3, 5, respectively. The transportation network is represented with a solid line, the communication network with a dashed line.

Regarding the cost functions, for each firm i, we set

$$c_e^i(t_e^i) = c_e(t_e^i) = t_e^i - \left(1 - \frac{1}{1+t_e^i}\right) \forall e, \tag{6.31a}$$

$$a^i(r^i) = a(r^i) = 2\left[r^i - \left(1 - \frac{1}{1+r^i}\right)\right]. \tag{6.31b}$$

For each market v, we consider the inverse demand function p_v to be affine and independent from the commodity sold at other markets. Specifically, for all v, $p_v(\sigma) = 10 - \sigma_v$. We assume that firm 2 bidirectionally communicates with firms 1 and 3, while

1 and 3 do not communicate, according to the communication matrix

$$T = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix},$$

which is primitive and doubly stochastic, hence satisfies Assumption 6.

We run Algorithm 5 with $\tau = 0.005$, $\nu = 10$ and initial conditions all equal to zero[9]. We use $\max\{\|x_{(k)} - x_{(k-1)}\|_\infty, \|\lambda_{(k)} - \lambda_{(k-1)}\|_\infty\} < 10^{-4}$ as stopping criterion. Figure 6.11 (top) reports the sales y^i for each firm in the 5 markets at the variational Nash equilibrium of $\mathcal{G}_\nu^{\text{CO}}$ (with $\nu = 10$), for the case when there are no coupling constraints (i.e., K is chosen so large that it has no effect). Figure 6.11 (bottom) reports how the equilibrium changes if we introduce the coupling constraint $[\sigma(x)]_3 \leq 1/3$, so that the total capacity of market 3 is 1. In both cases the $\mathcal{G}_\nu^{\text{CO}}$ (with $\nu = 10$) variational equilibrium is an ε_ν-Nash equilibrium for $\mathcal{G}^{\text{CO}}$, as by the second statement of Theorem 6. The value of ε_ν can be computed after convergence according to Definition 9. A more descriptive quantity is the relative maximum improvement $\hat{\varepsilon}_\nu$, defined as[10]

$$\hat{\varepsilon}_\nu := \max_{\substack{i \in \{1,\dots,M\} \\ x^i \in \mathcal{Q}^i(\bar{x}_\nu^{-i})}} \frac{J^i(\bar{x}_\nu^i, \sigma(\bar{x}_\nu)) - J^i(x^i, \frac{1}{M}x^i + \sum_{j \neq i} \frac{1}{M}\bar{x}_\nu^j)}{J^i(\bar{x}_\nu^i, \sigma(\bar{x}_\nu))}, \tag{6.32}$$

which equals 0.0014 (for the game without coupling constraint) and 0.0035 (for the game with coupling constraint).

Large network

As a more realistic example we consider the transportation network illustrated in Figure 6.13 which consists of $V = 43$ possible markets and $E = 51$ (bidirectional) edges connecting them. The network is taken from[11] the data set [Bri02], which provides also the Cartesian coordinates of the vertexes. We consider 5 firms that differ only for their locations ℓ^i, which are $\ell_1 = 37$, $\ell_2 = 20$, $\ell_3 = 11$, $\ell_4 = 6$, $\ell_5 = 35$ as indicated in Figure 6.13. Each firm has a production capacity of $\bar{r}^i = 10$, while we consider a capacity of 1.5 for each market (i.e. $K = 1.5/5$). The production cost is as in (6.31b) while the transportation cost for edge e is the same for each firm i and is

$$c_e^i(t_e^i) = c_e(t_e^i) = \rho_e \left(t_e^i - \left(1 - \frac{1}{1 + t_e^i} \right) \right),$$

[9] The values in (5.12) can be shown to be $\alpha_\nu = 4/(1+\tilde{r})^3 = 0.0185$, $L_\nu = \lambda_{\max}(H_{\text{blkd}}^\top[(I_M \otimes D)(T^\nu \otimes I_E) + \text{diag}(T^\nu) \otimes D^\top]H_{\text{blkd}}) = 9.9124$ and $\|A_\nu\| = 1$; then (5.12) reads $\tau < 1.8 \cdot 10^{-4}$. This is a conservative bound, we verified by simulations that the algorithm converges also for $\tau = 0.005$.

[10] Note that for any fixed $\bar{x}_\nu$, $\hat{\varepsilon}_\nu$ in (6.32) can be computed by solving the M optimization programs $\{\min_{x^i \in \mathcal{Q}^i(\bar{x}_\nu^{-i})} J^i(x^i, \frac{1}{M}x^i + \sum_{j \neq i} \frac{1}{M}\bar{x}_\nu^j)\}_{i=1}^M$.

[11] As in Section 6.2, we consider a subgraph of the original data set for computational tractability.

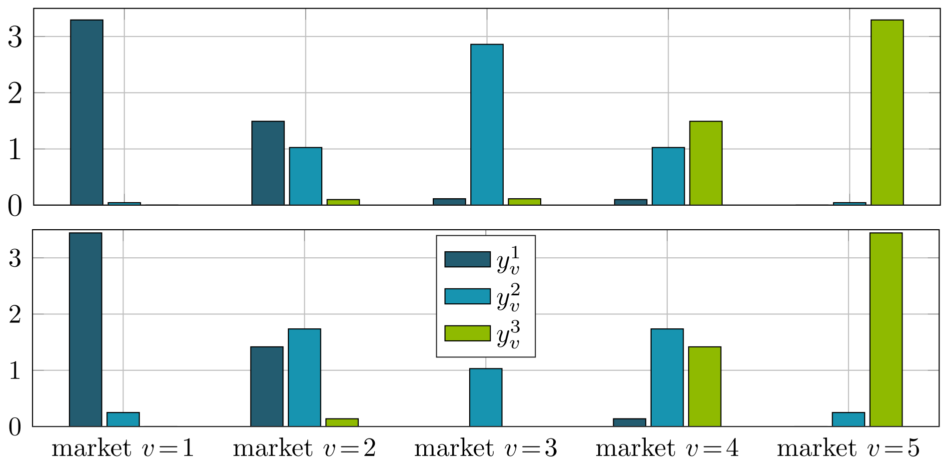


Figure 6.11: Production per firm and market without coupling constraint (top) and with coupling constraint $[\sigma^i(x)]_3 \leq 1/3$ (bottom). In both cases the total production at the equilibrium is $r^i = \bar{r}^i = 5$ for all i. Both simulations are obtained with $\nu = 10$ communications.

where $\rho_e \in]0,1]$ is the normalized[12] length of road e. The inverse demand function p is affine, i.e. $p(\sigma) = 10 \cdot \mathbb{1}_{45} - D\sigma$ and it encodes intra-market competition via the matrix D whose component in position (h,k) is $[D]_{h,k} = 1$ if $h = k$, $[D]_{h,k} = 0.3(1-\rho_e)$, if there is a road $e = (h,k)$ between markets h and k, while $[D]_{h,k} = 0$ otherwise. In words, the price p_v at market v not only decreases when more commodity is sold at v, but also when more commodity is sold at the neighboring markets, with physically close markets being more influential. We verified numerically that $D \succeq 0$. We use the communication matrix T that corresponds to a symmetric ring, i.e.,

$$T = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \end{bmatrix},$$

which satisfies Assumption 6. We run Algorithm 5 with $\tau = 0.05$, initial conditions all equal to zero and different values of ν[13]. As for the small network, we use $\max\{\|x_{(k)} - x_{(k-1)}\|_\infty, \|\lambda_{(k)} - \lambda_{(k-1)}\|_\infty\} < 10^{-4}$ as stopping criterion. We consider even values of ν between 2 and 20. For each ν we run Algorithm 5 and find the variational Nash equilibrium of $\mathcal{G}_\nu^{\text{CO}}$, which is an ε_ν-Nash equilibrium for $\mathcal{G}^{\text{CO}}$, as by the second statement

[12] The normalized length of a road is defined as the absolute length divided by the maximum length road in the network.

[13] The values in (5.12) can be shown to be $\alpha_\nu = 4/(1+\tilde{r})^3 = 0.003$, $L_\nu = \lambda_{\max}(H_{\text{blkd}}^\top[(I_M \otimes D)(T^\nu \otimes I_E) + \text{diag}(T^\nu) \otimes D^\top]H_{\text{blkd}}) = 12.89$ and $\|A_\nu\| = 1$; then (5.12) reads $\tau < 1.8 \cdot 10^{-5}$. This is a conservative bound, we verified by simulations that the algorithm converges also for $\tau = 0.05$.

of Theorem 6. After convergence ε_ν can be computed according to Definition 9 and $\hat{\varepsilon}_\nu$ according to (6.32). Figure 6.12 (top) reports the value of $\hat{\varepsilon}_\nu$ as function of ν, thus numerically verifying the second statement of Theorem 6. Figure 6.12 (bottom) reports the value of $\|\bar{x}_\nu - \bar{x}\|_2$ as function of ν, thus numerically verifying the first statement (5.5) of Theorem 6. In Figure 6.13 we illustrate the variational Nash equilibrium of $\mathcal{G}^{\mathrm{CO}}$ obtained by setting $\nu = 1$ and $T = \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top$. We note that each firm is the only seller at the location where it produces, and more in general firms tend to sell close to their production location, as expected.

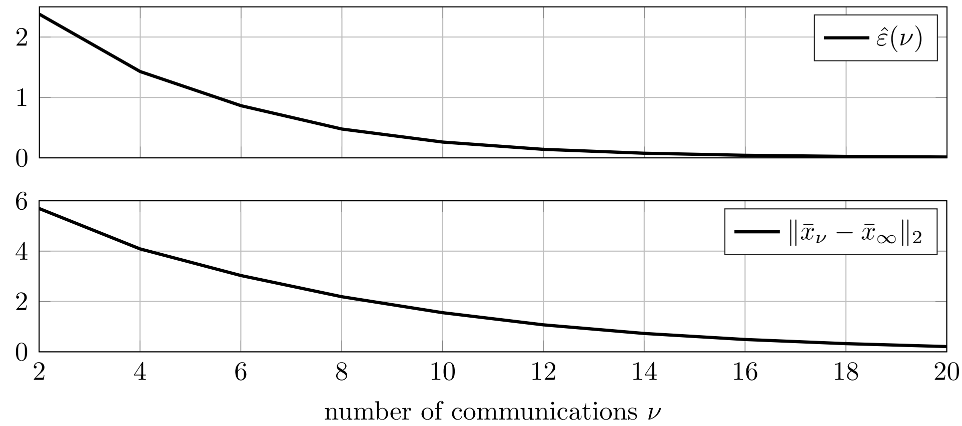


Figure 6.12: Relative maximum cost improvement $\hat{\varepsilon}_\nu$ in (6.32) as a function of ν.

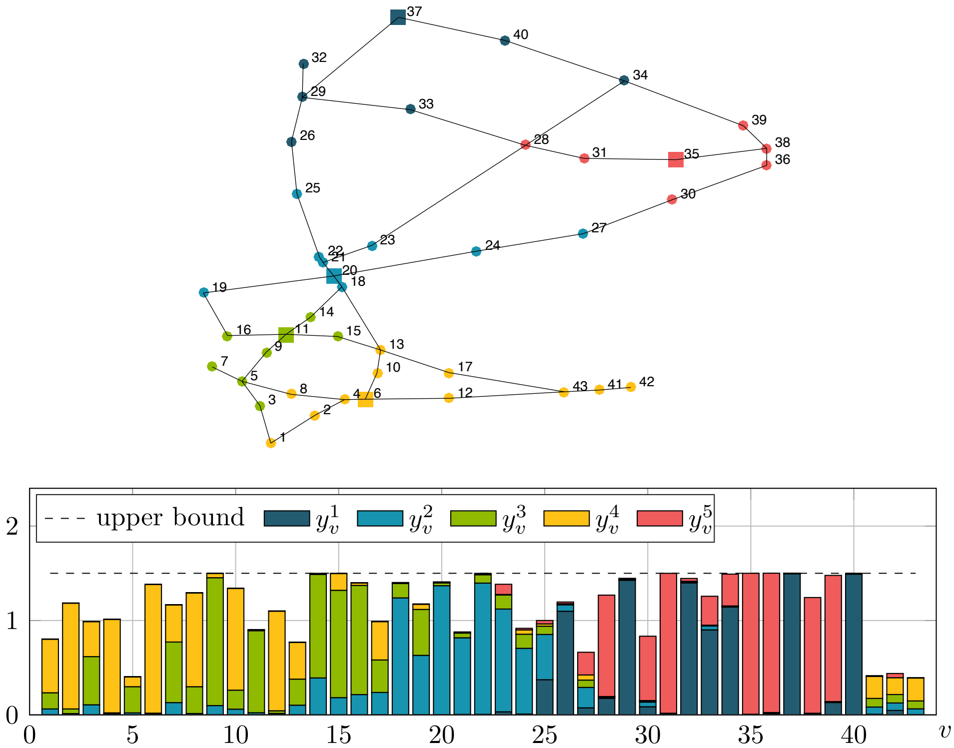


Figure 6.13: Variational Nash equilibrium of $\mathcal{G}^{\mathrm{CO}}$ for the large network, computed by Algorithm 1 for $\nu = 1$ and $T = \frac{1}{M}\mathbb{1}_M\mathbb{1}_M^\top$. In the top plot, each market takes the color of the firm that sells the most commodity in that market. The production locations of the firms are denoted by squares. The bottom plot reports y_v^i for each agent i and market v.

CHAPTER 7

The Wardrop equilibrium with inertia

7.1 Definition and examples

While the previous chapters focused on equilibria defined in the strategy space x, we here consider an equilibrium in $\sigma(x)$, analogously to what done in (3.10) for the Wardrop equilibrium of a parallel road network. As we only deal with the aggregate $\sigma(x)$, we discard the dependence from the strategies x and refer directly to σ. We start by introducing the definition of parallel Wardrop, which coincides with conditions (3.10) relative to the Wardrop equilibrium of parallel roads.

Definition 11 (Parallel Wardrop equilibrium)**.** Given utilities $\{u_j : \mathbb{R}_{\geq 0} \to \mathbb{R}\}_{j=1}^n$ and mass $\gamma > 0$, the vector $\bar{\sigma} \in \mathbb{R}^n$ is a *parallel Wardrop equilibrium* if $\bar{\sigma} \geq \mathbb{0}_n$, $\mathbb{1}_n^\top \bar{\sigma} = \gamma$ and for all $j \in \{1, \dots, n\}$

$$\bar{\sigma}_j > 0 \Rightarrow u_j(\sigma_j) \geq u_h(\sigma_h), \quad \forall\, h \in \{1, \dots, n\}. \tag{7.1}$$

□

Compared to conditions (3.10), Definition 11 is based on maximization of the utilities $\{u_j\}_{j=1}^n$ rather than minimization of the travel times $\{t_j\}_{j=1}^n$ (to establish the equivalence it suffices to set $u_j = -t_j$) and it assumes that utility u_j is a sole function of σ_j rather than being a function of the entire vector $\sigma = [\sigma_j]_{j=1}^n$.

The fact that in Definition 11 only actions with maximum utility are used hinges upon the ability of switching between actions without incurring any cost. The next Definition 11 accounts for the fact that in many practical situations such switches do not come for free.

Definition 12 (Inertial Wardrop equilibrium)**.** Given utilities $\{u_j : \mathbb{R}_{\geq 0} \to \mathbb{R}\}_{j=1}^n$, inertia coefficients $\{c_{jh} \geq 0\}_{j,h=1}^n$ and mass $\gamma > 0$, the vector $\bar{\sigma} \in \mathbb{R}^n$ is an *inertial Wardrop equilibrium* if $\bar{\sigma} \geq \mathbb{0}_n$, $\mathbb{1}_n^\top \bar{\sigma} = \gamma$ and for all $j \in \{1, \dots, n\}$

$$\bar{\sigma}_j > 0 \Rightarrow u_j(\sigma_j) \geq u_h(\sigma_h) - c_{jh}, \quad \forall\, h \in \{1, \dots, n\}. \tag{7.2}$$

□

To comment on the definition of inertial Wardrop (which we also refer to as Wardrop with inertia) we interpret σ_j as the sum of infinitesimal agents who are choosing action j. The no-improvement condition (7.2) states that the infinitesimal agents at action j do not have any incentive in switching to action h, considering the difference in utilities and the cost of switching (or inertia coefficient) c_{jh}. The presence of the coefficients $\{c_{jh}\}_{h,j=1}^{n}$ constitutes the only difference between the standard Definition 11 and the novel Definition 12. The coefficients can model different phenomena, such as

- an effective cost or fee that the agents incur for switching action;
- the attitude of the agents to adhere to their habits, or their reluctance to trying something different;
- the lack of information about other options.

Note that the no-improvement condition (7.2) does not impose anything on empty actions, i.e., actions that are not taken by any agent. In other words, the utility of an empty action can be arbitrarily bad and the configuration $\bar{\sigma}$ still be an equilibrium.

We conveniently define the simplex and the positive simplex as

$$\mathcal{S} := \left\{ \sigma \in \mathbb{R}^n_{\geq 0} \mid \sum_{j=1}^{n} \sigma_j = \gamma \right\} \subset \mathbb{R}^n, \qquad \mathcal{S}_{>0} := \left\{ \sigma \in \mathbb{R}^n_{>0} \mid \sum_{j=1}^{n} \sigma_j = \gamma \right\} \subset \mathbb{R}^n.$$

Given $\sigma \in \mathcal{S}$, we say that the infinitesimal agents at action j are jealous of action h (or that action h is attractive for agents at j) if $u_j(\sigma_j) < u_h(\sigma_h) - c_{jh}$. The next example provides some intuition about the equilibrium set.

Example 4. Let us consider three possible actions (that is, $n = 3$) with utilities and inertia coefficients

$$\begin{aligned} u_1(\sigma_1) &= 1.2 - \sigma_1 \\ u_2(\sigma_2) &= 1.2 - \sigma_2 \\ u_3(\sigma_3) &= 1 - \sigma_3, \end{aligned} \qquad C = \begin{bmatrix} 0 & 0.2 & 0.3 \\ 1 & 0 & 0.8 \\ 0.1 & 1.2 & 0 \end{bmatrix}, \tag{7.3}$$

where the entry (j,h) of C equals c_{jh}. We take $\gamma = 1$, so that $\sigma_3 = 1 - \sigma_1 - \sigma_2$. We explicitly compute the six conditions (7.2):

$$\begin{array}{llllll}
\sigma_1 > 0 & \Rightarrow & 1.2 - \sigma_1 + 0.2 \geq 1.2 - \sigma_2 & \Leftrightarrow & \boxed{\sigma_2 \geq \sigma_1 - 0.2} & \text{(7.4a)} \\
\sigma_1 > 0 & \Rightarrow & 1.2 - \sigma_1 + 0.3 \geq 1 - (1 - \sigma_1 - \sigma_2) & \Leftrightarrow & \boxed{\sigma_2 \leq -2\sigma_1 + 1.5} & \text{(7.4b)} \\
\sigma_2 > 0 & \Rightarrow & 1.2 - \sigma_2 + 1 \geq 1.2 - \sigma_1 & \Leftrightarrow & \sigma_2 \leq \sigma_1 + 1 & \text{(7.4c)} \\
\sigma_2 > 0 & \Rightarrow & 1.2 - \sigma_2 + 0.8 \geq 1 - (1 - \sigma_1 - \sigma_2) & \Leftrightarrow & \sigma_2 \leq -0.5\sigma_1 + 1 & \text{(7.4d)} \\
\sigma_3 > 0 & \Rightarrow & 1 - (1 - \sigma_1 - \sigma_2) + 0.1 \geq 1.2 - \sigma_1 & \Leftrightarrow & \boxed{\sigma_2 \geq -2\sigma_2 + 1.1} & \text{(7.4e)} \\
\sigma_3 > 0 & \Rightarrow & 1 - (1 - \sigma_1 - \sigma_2) + 1.2 \geq 1.2 - \sigma_2 & \Leftrightarrow & 2\sigma_2 \geq -\sigma_1, & \text{(7.4f)}
\end{array}$$

where we strike through the inequalities which are implied by $\sigma \in \mathcal{S}$, and we color code the remaining three according to Figure 7.1, which reports the solution to (7.4) (i.e, the equilibrium set) computed numerically. The figure plots only the variables σ_1 and σ_2, because $\sigma_3 = 1 - \sigma_1 - \sigma_2$.

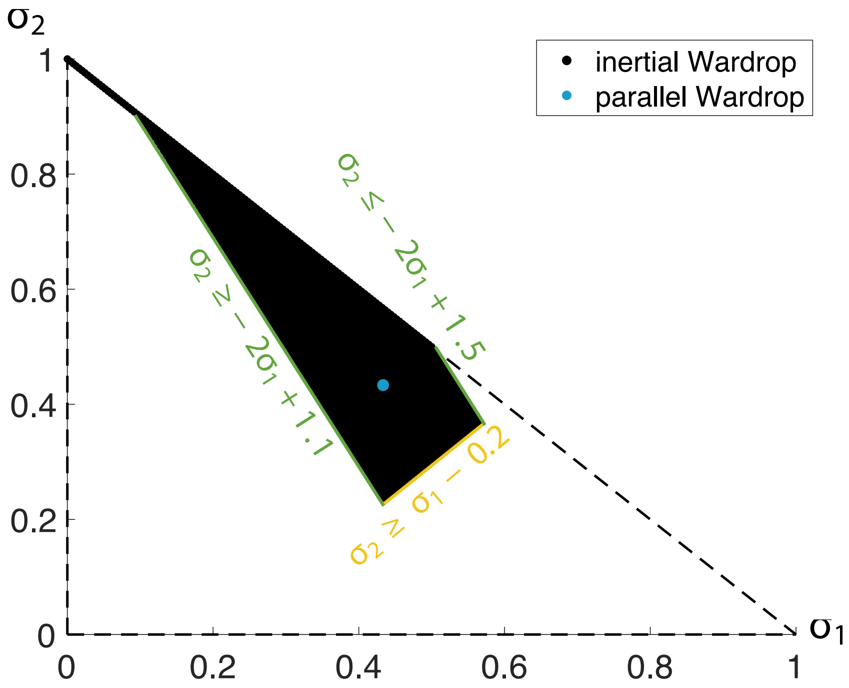


Figure 7.1: In black the equilibrium set for Example 4 obtained by gridding the space and verifying for every point whether it satisfies Definition 12 or not. In dashed line we indicate the simplex boundary, in green and yellow inequalities (7.4). The set is visibly non-convex because it includes the segment $[0.1, 0.9] - [0, 1]$. The point in cyan is the parallel Wardrop $[0.4, 0.4]$, which satisfies condition (7.1).

The first observation is that the inertial Wardrop equilibrium set is visibly not a singleton; this marks a difference with Chapters 3-6, where the focus is mostly on unique Nash or Wardrop equilibria. The lack of uniqueness is due to the positivity of the inertia coefficients c_{jh}. Indeed, if $c_{jh} = 0$ for all j, h, then the inertial Wardrop of Definition 12 coincides with the parallel Wardrop of Definition 11, which in Figure 7.1 is marked in cyan and is unique, as explained in the following Section 7.2.

As a second observation, we point out the non-convexity of the inertial Wardrop equilibrium set, which is due to the presence of the segment $[0.1, 0.9] - [0, 1]$. The points of the segment belong to the equilibrium set even though they do not satisfy the green inequality in (7.4e). Indeed, this is the case because (7.4e) enforces the green inequality only when $\sigma_3 > 0$, whereas $\sigma_3 = 0$ in the segment. Non-convexity is never an issue in Chapters 3-6. □

Example 5. Let us change utilities and inertia coefficients compared to (7.3)

$$\begin{aligned} u_1(\sigma_1) &= -2.5\sigma_1^2 - 0.51\sigma_1 \\ u_2(\sigma_2) &= -3.75\sigma_2^2 - 0.33\sigma_2 \\ u_3(\sigma_3) &= -0.5\sigma_3^2 - 0.42\sigma_3. \end{aligned} \qquad C = \begin{bmatrix} 0 & 0.4 & 0.9 \\ 0.8 & 0 & 1 \\ 0.3 & 0.6 & 0 \end{bmatrix}, \tag{7.5}$$

The total mass is again $\gamma = 1$ and we report in Figure 7.2 the inertial Wardrop equilibrium set.

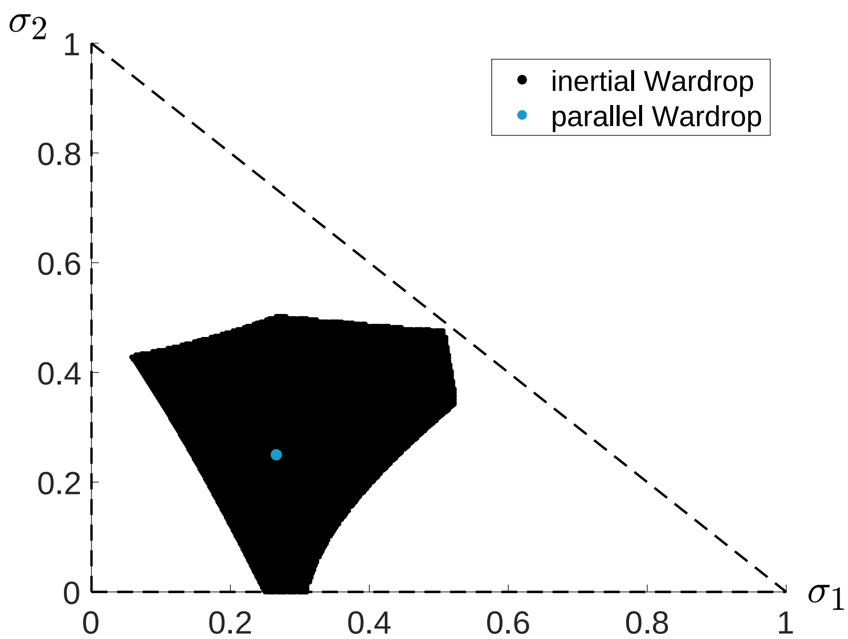


Figure 7.2: In black the equilibrium set for Example 5 obtained by gridding the space and verifying for every point whether it satisfies Definition 12 or not. In dashed line we indicate the simplex boundary. The set is visibly non-convex. The point in cyan is the parallel Wardrop of Definition 11.

Contrary to Example 4, non-convexity of the equilibrium set arises because a point $\sigma \in \mathcal{S}_{>0}$ is an inertial Wardrop if and only if it satisfies the system

$$\begin{aligned} u_1(\sigma_1) &\le u_2(\sigma_2) - c_{12}, & u_1(\sigma_1) &\le u_3(\sigma_3) - c_{13}, \\ u_2(\sigma_2) &\le u_1(\sigma_1) - c_{21}, & u_2(\sigma_2) &\le u_3(\sigma_3) - c_{23}, \\ u_3(\sigma_3) &\le u_1(\sigma_1) - c_{31}, & u_3(\sigma_3) &\le u_2(\sigma_2) - c_{32}. \end{aligned} \tag{7.6}$$

If the utilities were affine as in Example 4, then the solution set of (7.6) would be convex. Instead, with non-affine utilities it is not guaranteed to be convex, and indeed it is not with the specific choice (7.5). □

7.2 Relation with the parallel Wardrop equilibrium

In this section we propose an algorithm for finding an inertial Wardrop and highlight its drawbacks.

Lemma 14. A point $\bar{\sigma} \in \mathbb{R}^n$ is a parallel Wardrop equilibrium if and only if it is a solution of VI($\mathcal{S}$,$-u$), where $u = [u_j]_{j=1}^n$. □

Proof. The proof can be conducted by exploiting the derivations of Example 2, but we report it here in full for completeness. By Proposition 4, $\bar{\sigma}$ is a solution of VI($\mathcal{S}$,$-u$) if and only if it solves its KKT system:

$$-u(\sigma) - \lambda + \mu \mathbb{1}_n = \mathbb{0}_n \tag{7.7a}$$
$$\mathbb{0}_n \leq \lambda \perp \sigma \geq \mathbb{0}_n, \tag{7.7b}$$
$$\mathbb{1}_n^\top \sigma = \gamma, \tag{7.7c}$$

where $\lambda \in \mathbb{R}^n$ is the dual variable corresponding to the non-negativity constraint and $\mu \in \mathbb{R}$ is the dual variable corresponding to the constraint $\mathbb{1}_n^\top \sigma = \gamma$. By substituting (7.7a) into (7.7b), the system (7.7) can be simplified into

$$\mathbb{0}_n \leq -u(\sigma) + \mu \mathbb{1}_n \perp \sigma \geq \mathbb{0}_n, \tag{7.8a}$$
$$\mathbb{1}_n^\top \sigma = \gamma. \tag{7.8b}$$

We now argue that it must hold

$$\mu = u_{\max}(\sigma) := \max_{j \in \{1,\dots,n\}} u_j(\sigma_j).$$

Indeed, if $\mu > u_{\max}(\sigma)$ then $-u(\sigma) + \mu \mathbb{1}_n > \mathbb{0}_n$, hence by the orthogonality condition (7.8a) $\sigma = \mathbb{0}_n$ which violates (7.8b). If instead $\mu < u_{\max}(\sigma)$ then $-u(\sigma) + \mu \mathbb{1}_n \geq \mathbb{0}_n$ does not hold. We can conclude that (7.8) reads as

$$\mathbb{0}_n \leq -u(\sigma) + u_{\max}(\sigma) \mathbb{1}_n \perp \sigma \geq \mathbb{0}_n,$$
$$\mathbb{1}_n^\top \sigma = \gamma_{\text{avg}},$$

which is equivalent to

$$\begin{aligned} &\sigma \geq \mathbb{0}_n, \qquad \mathbb{1}_n^\top \sigma = \gamma \\ &\sigma_j > 0 \Rightarrow u_j(\sigma_j) = u_{\max}(\sigma), \quad \forall\, j \in \{1, \dots, n\}, \end{aligned} \tag{7.10}$$

and conditions (7.10) coincide with Definition 11. □

Lemma 15. Every parallel Wardrop equilibrium is an inertial Wardrop equilibrium.

Proof. It follows directly from Definitions 11 and 12, since condition (7.1) implies condition (7.2), as $c_{jh} \geq 0$ for all $j, h \in \{1, \dots, n\}$. □

We cast in the following the projection algorithm of Section 2.3 to VI($\mathcal{S}$,$-u$), where $\sigma(k)$ indicates the iterate k of the algorithm.

Algorithm 6: Projection algorithm

Initialization	$\tau > 0$, $k = 0$, $\sigma(0) \in \mathcal{S}$
Iterate	$\sigma(k+1) = \operatorname*{Proj}_{\mathcal{S}} [\sigma(k) + \tau u(\sigma(k))]$
	$k \leftarrow k+1$

Lemma 16. If u_j is non-increasing in $[0, \gamma]$ and L-Lipschitz for all j, then Algorithm 6 converges to a parallel Wardrop equilibrium for $\tau < 2/L$.

Proof. Since u_j is non-increasing for all j, then

$$(-u(\sigma^1) + u(\sigma^2))^\top(\sigma^1 - \sigma^2) = \sum_{j=1}^{n} \underbrace{(-u_j(\sigma_j^1) + u_j(\sigma_j^2))(\sigma_j^1 - \sigma_j^2)}_{\geq 0} \geq 0, \quad \forall \sigma^1, \sigma^2 \in \mathcal{S},$$

hence $-u$ is monotone by Definition 3. As u_j is Lipschitz, then it is continuous, hence by the fundamental theorem of calculus it admits a primitive U_j. It follows that $\nabla_\sigma \left(-\sum_{j=1}^n U_j(\sigma_j)\right) = -u(\sigma)$, hence $-u$ is a gradient operator. A solution of VI($\mathcal{S}$,$-u$) exists by Proposition 6, because $\mathcal{S}$ is convex, compact and $-u$ is continuous. Thus all the assumptions of Corollary 1 are met and its statement concludes the proof. □

Corollary 5. Under the assumptions of Lemma 16, Algorithm 6 converges to an inertial Wardrop equilibrium.

Proof. The statement is a straightforward consequence of Lemmas 15 and 16. □

Example 4 (continued). Thanks to Lemma 14, we can show uniqueness of the parallel Wardrop equilibrium in Example 4. Indeed, each u_j in (7.3) is strictly decreasing, hence the operator $-u$ is strongly monotone in the set $\mathcal{S}$. Then VI($\mathcal{S}$,$-u$) admits a unique solution by Proposition 7. The following shows that the solution is $[\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3] = [0.4, 0.4, 0.2]$.

$$\begin{bmatrix} -u_1(\bar{\sigma}_1) \\ -u_2(\bar{\sigma}_2) \\ -u_3(\bar{\sigma}_3) \end{bmatrix}^\top \left(\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} - \begin{bmatrix} \bar{\sigma}_1 \\ \bar{\sigma}_2 \\ \bar{\sigma}_3 \end{bmatrix} \right) = \begin{bmatrix} -0.8 \\ -0.8 \\ -0.8 \end{bmatrix}^\top \left(\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} - \begin{bmatrix} \bar{\sigma}_1 \\ \bar{\sigma}_2 \\ \bar{\sigma}_3 \end{bmatrix} \right) =$$
$$(-0.8) \left(\sum_{j=1}^{3} \sigma_j - \sum_{j=1}^{3} \bar{\sigma}_j \right) = (-0.8)(1-1) = 0, \quad \forall\, \sigma \in \mathcal{S}.$$

Lemma 15 reflects in the fact that the cyan point corresponding to the parallel Wardrop in Figure 7.1 is within the inertial Wardrop equilibrium set.

We now analyze Algorithm 6 on this simple example to highlight its drawbacks. By Corollary 5, convergence of Algorithm 6 to the parallel Wardrop is guaranteed for $\tau < 2$, because the Lipschitz constant L of the utilities in (7.4) equals 1. If we choose $\tau = 1$, the first iteration of Algorithm 6 becomes

$$\sigma(1) = \operatorname*{Proj}_{\mathcal{S}} \Big[\sigma(0) + u(\sigma(0))\Big] = \operatorname*{Proj}_{\mathcal{S}} \left[\sigma(0) + \begin{bmatrix}1.2\\1.2\\1\end{bmatrix} - \sigma(0)\right] = \operatorname*{Proj}_{\mathcal{S}} \left[\begin{bmatrix}1.2\\1.2\\1\end{bmatrix}\right] = \begin{bmatrix}0.4\\0.4\\0.2\end{bmatrix},$$

so one iteration is enough to converge to the parallel Wardrop, for any initial condition $\sigma(0)$.

Firstly, we consider as initial condition $\sigma(0) = [0.4; 0.2; 0.4]$, which is not an inertial Wardrop, because $\sigma_3(0) > 0$ and $u_3(\sigma_3(0)) = 1 - 0.4 = 0.6 < 0.7 = 0.8 - 0.1 = u_1(\sigma_1(0)) - c_{31}$. In other words, the agents at action 3 are jealous of the agents at action 1. On the contrary, agents at 3 are not jealous of agents at 2, because $u_3(\sigma_3(0)) = 0.6 \geq -0.2 = u_2(\sigma_2(0)) - c_{32}$. The first iteration of Algorithm 6 consists in 0.2 units of agents switching from action 3 to action 2, even though such switch is detrimental for the agents performing it. Indeed, the agents at 3 would rather prefer to switch to action 1.

Secondly, we consider the initial condition $\sigma(0) = [0.4; 0.3; 0.3]$. It can be verified that $\sigma(0)$ belongs to the inertial Wardrop equilibrium set. Nonetheless, in the first iteration of Algorithm 6, 0.1 units of agents move from action 3 to action 2. □

We are now ready to list the two drawbacks of Algorithm 6.

1. The agents are forced to switch action even when such switch is detrimental, as highlighted in Example 4. In particular, they might be forced to switch even if already at an inertial Wardrop equilibrium.

2. To perform the projection operation of Algorithm 6, the agents need to know not only the utility $u(\sigma(k))$ of all actions, but also the amount of agents $\sigma(k)$ performing such actions.

Section 7.4 proposes an algorithm that overcomes these two drawbacks. To streamline the presentation, we first establish the connection between inertial Wardrop and variational inequality.

7.3 Variational inequality reformulation of the inertial Wardrop equilibrium

In this section we show that the set of inertial Wardrop equilibria coincides with the solution set of a certain variational inequality, which is different from the one relative to the parallel Wardrop. We then study monotonicity of the VI operator.

Let us define the operator $F : \mathcal{S} \to \mathbb{R}^n_{\geq 0}$ as follows

$$\begin{aligned} F(\sigma) &= [F_j(\sigma)]_{j=1}^n, \\ F_j(\sigma) &= \max_{h\in\{1,\dots,n\}} (u_h(\sigma_h) - u_j(\sigma_j) - c_{jh}), \end{aligned} \tag{7.11}$$

where we use the convention that $c_{jj} = 0$ for all $j \in \{1, \dots, n\}$, as in (7.3). Then one of the elements of $\{u_h(\sigma_h) - u_j(\sigma_j) - c_{jh}\}_{h=1}^n$ is $u_j(\sigma_j) - u_j(\sigma_j) - c_{jj} = 0$, hence for all $j \in \{1, \dots, n\}$ it holds $F_j(\sigma) \geq 0$ for all $\sigma \in \mathcal{S}$. Moreover, by Definition 12 the agents at j find another action attractive if and only if $u_h(\sigma_h) - u_j(\sigma_j) - c_{jh} > 0$; in other words, $F_j(\sigma) = 0$ if and only if the agents at j are not jealous of any other action. The next lemma clarifies that $F(\sigma) > \mathbb{0}_n$ is not possible.

Lemma 17. For each $\sigma \in \mathcal{S}$ there exists $j^\star$ such that $F_{j^\star}(\sigma) = 0$.

Proof. Take $j^\star \in \operatorname*{argmax}_{j\in\{1,\dots,n\}} u_j(\sigma_j)$. Then $F_{j^\star}(\sigma) = 0$ by its definition (7.11). □

Theorem 9. A vector $\bar{\sigma} \in \mathbb{R}^n$ is an inertial Wardrop equilibrium if and only if it is a solution of $\mathrm{VI}(\mathcal{S}, F)$.

Proof. The proof consists in showing that the KKT system of $\mathrm{VI}(\mathcal{S}, F)$ is equivalent to Definition 12 of inertial Wardrop. Since the set $\mathcal{S}$ satisfies Slater's constraint qualification, by Proposition 4 $\mathrm{VI}(\mathcal{S}, F)$ is equivalent to its KKT system

$$F(\sigma) + \mu\mathbb{1}_n - \lambda = \mathbb{0}_n \tag{7.12a}$$
$$\sigma \geq \mathbb{0}_n \tag{7.12b}$$
$$\mathbb{1}_n^\top \sigma = \gamma \tag{7.12c}$$
$$\lambda \geq \mathbb{0}_n \tag{7.12d}$$
$$\lambda^\top \sigma = 0, \tag{7.12e}$$

where $\mu \in \mathbb{R}$ is the dual variable corresponding to the constraint $\mathbb{1}_n^\top \sigma = \gamma$ and $\lambda \in \mathbb{R}^n$ is the dual variable corresponding to the constraint $\sigma \geq \mathbb{0}_n$. The system (7.12) can be compactly rewritten as

$$\mathbb{0}_n \leq \mu\mathbb{1}_n + F(\sigma) \perp \sigma \geq \mathbb{0}_n, \tag{7.13a}$$
$$\mathbb{1}_n^\top \sigma = \gamma. \tag{7.13b}$$

Lemma 17 ensures the existence of $j \in \{1, \dots, n\}$ such that $F_j(\sigma) = 0$. Then $\mu < 0$ is not possible, otherwise the non-negativity condition on $\mu\mathbb{1}_n + F(\sigma)$ is violated. Moreover, since $F(s) \geq \mathbb{0}_n$, $\mu > 0$ is not possible, as by (7.13a) this would imply $\sigma = \mathbb{0}_n$ thus violating (7.13b). We can conclude that $\mu = 0$ and (7.13) becomes

$$\mathbb{0}_n \leq F(\sigma) \perp \sigma \geq \mathbb{0}_n, \tag{7.14a}$$
$$\mathbb{1}_n^\top \sigma = \gamma. \tag{7.14b}$$

The system (7.14) is equivalent to

$$\sigma \in \mathcal{S}, \text{ and}$$
$$\sigma_j > 0 \underset{(7.14a)}{\Rightarrow} u_j(\sigma_j) \geq u_h(\sigma_h) - c_{jh}, \ \forall\, j, h \in \{1, \ldots, n\}.$$

which coincides with Definition 12. □

Corollary 6. For any set of non-negative inertia coefficients $\{c_{jh} \geq 0\}_{j,h=1}^n$, continuous utility functions $\{u_i : \mathbb{R}_{\geq 0} \to \mathbb{R}\}_{i=1}^n$ and total mass $\gamma > 0$, there exists an inertial Wardrop equilibrium.

Proof. It follows from Theorem 9 and Proposition 6 on existence of VI solutions, because F is continuous as each of its components is the point-wise maximum of continuous functions. □

Absence of monotonicity

If $\text{VI}(\mathcal{S}, F)$ exhibits any of the monotonicity properties of Table 2.3, then convergence of the projection or the extragradient algorithm would be guaranteed. However, the following classical result about variational inequality solutions clarifies that monotonicity cannot be always guaranteed.

Proposition 18 ([FP03, Theorem 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 $\text{VI}(\mathcal{X}, F)$ is convex. □

Since the inertial Wardrop equilibrium set in Figure 7.1 is not convex, then the corresponding variational inequality operator F cannot be monotone. In the following we generalize such observation by providing a weak sufficient condition for absence of monotonicity. To this end, we introduce a generalization of Proposition 8, because therein the VI operator is assumed to be continuously differentiable, whereas F is not.

Proposition 19. [DTL96, Proposition 2.1] An operator F is monotone in $\mathcal{X} \subseteq \mathbb{R}^n$ if and only if for every $x \in \mathcal{X}$ each generalized Jacobian $\phi \in \partial F(\sigma)$ is positive semidefinite. □

The definition of generalized Jacobian $\partial F(\sigma)$ can be found in [Cla90, Definition 2.6.1]; we do not report it here because all we need in the proof of the following Theorem 10 is that if F is differentiable at σ, then $\partial F(\sigma) = \{\nabla_\sigma F(\sigma)\}$. In words, the generalized Jacobian coincides with the Jacobian. The following Theorem 10 clarifies that in presence of strictly decreasing utilities the inertial Wardrop operator F in (7.11) is never monotone, except for the degenerate case in which the entire set $\mathcal{S}$ is an inertial Wardrop equilibrium.

Theorem 10. Assume that in $[0,\gamma]$ the function u_j is strictly decreasing and Lipschitz for all $j \in \{1,\dots,n\}$, that $c_{jh} \geq 0$ for all $j,h \in \{1,\dots,n\}$ and that there exists a point $\hat{\sigma} \in \mathcal{S}$ which is not an inertial Wardrop equilibrium. Then F is not monotone in $\mathcal{S}$.

Proof. The proof is composed by four parts.

1) We first show that there exists $\tilde{\sigma} \in \mathcal{S}_{>0}$ such that $\tilde{\sigma}$ is not an inertial Wardrop.
For the sake of contradiction, assume that each $\sigma \in \mathcal{S}_{>0}$ is an inertial Wardrop. Since $\hat{\sigma}$ belongs to the closure of $\mathcal{S}_{>0}$, we can construct a sequence $(\sigma(m))_{m=1}^{\infty} \in \mathcal{S}_{>0}$ such that $\lim_{m\to\infty} \sigma(m) = \hat{\sigma}$. Since each $\sigma(m)$ is an inertial Wardrop and it is positive, then for all j,h it holds $u_j(\sigma_j(m)) \geq u_h(\sigma_h(m)) - c_{jh}$. Taking the limit and exploiting continuity of $\{u_j\}_{j=1}^n$ we obtain

$$\begin{aligned} &\lim_{m\to\infty} u_j(\sigma_j(m)) \geq \lim_{m\to\infty} u_h(\sigma_h(m)) - c_{jh}, \quad \forall\, j,h \in \{1,\dots,n\} \\ &\Leftrightarrow u_j(\hat{\sigma}_j) \geq u_h(\hat{\sigma}_h) - c_{jh}, \quad \forall\, j,h \in \{1,\dots,n\}, \end{aligned} \tag{7.15}$$

hence $\hat{\sigma}$ is an inertial Wardrop equilibrium, against the assumption.

2) After establishing the existence of $\tilde{\sigma} \in \mathcal{S}_{>0}$ which is not an inertial Wardrop, we now show that there exists an open ball $\mathcal{B}_{\tilde{\varepsilon}}(\tilde{\sigma})$ centered around $\tilde{\sigma}$ of radius $\tilde{\varepsilon} > 0$ such that none of the points in $\mathcal{B}_{\tilde{\varepsilon}}(\tilde{\sigma}) \cap \mathcal{S}_{>0}$ is an inertial Wardrop. Let us reason again for the sake of contradiction. If for each $\varepsilon > 0$ there exists an inertial Wardrop in $\mathcal{B}_{\varepsilon}(\tilde{\sigma}) \cap \mathcal{S}_{>0}$, then we can construct a sequence of inertial Wardrop equilibria converging to $\tilde{\sigma}$. With the same continuity argument used in (7.15), we can conclude that $\tilde{\sigma}$ is an inertial Wardrop, which is false by assumption. This demonstrates the existence of $\tilde{\varepsilon} > 0$ such that none of the points in $\mathcal{B}_{\tilde{\varepsilon}}(\tilde{\sigma}) \cap \mathcal{S}_{>0}$ is an inertial Wardrop. By Rademacher's theorem [AFP00, Theorem 2.14], Lipschitzianity of $\{u_j\}_{j=1}^n$ guarantees[1] existence of $\sigma^\star \in \mathcal{B}_{\tilde{\varepsilon}}(\tilde{\sigma}) \cap \mathcal{S}_{>0}$ such that F is differentiable at $\sigma^\star$.

3) The previous part guarantees differentiability of F at a point $\sigma^\star \in \mathcal{S}_{>0}$ which is not an inertial Wardrop. This third part is dedicated to showing that there exist $j^\star, h^\star \in \{1,\dots,n\}$ such that $j^\star \in \mathcal{A}(h^\star,\sigma^\star)$ and $\mathcal{A}(j^\star,\sigma^\star) = \{j^\star\}$, where we denote

$$\mathcal{A}(k,\sigma) := \operatorname*{argmax}_{\ell \in \{1,\dots,n\}} \{u_\ell(\sigma_\ell) - u_k(\sigma_k) - c_{k\ell}\}. \tag{7.16}$$

Since $\sigma^\star$ is not an inertial Wardrop, then there exist ℓ_1, ℓ_2 such that

$$u_{\ell_1}(\sigma^\star_{\ell_1}) < u_{\ell_2}(\sigma^\star_{\ell_2}) - c_{\ell_1\ell_2}. \tag{7.17}$$

Condition (7.17) is equivalent to $\ell_2 \in \mathcal{A}(\ell_1,\sigma^\star)$ and $\ell_1 \notin \mathcal{A}(\ell_1,\sigma^\star)$. If $\mathcal{A}(\ell_2,\sigma^\star) = \{\ell_2\}$ then the statement is proven with $h^\star = \ell_1, j^\star = \ell_2$, otherwise there exists $\ell_3 \in$

[1]Radamacher's theorem assumes F to be defined on an open subset of $\mathbb{R}^n$, but $\mathcal{S}_{>0}$ is not open in $\mathbb{R}^n$. Indeed, one just needs to define F on the $n-1$ dimensional open set $\{\sigma \in \mathbb{R}^{n-1}_{\geq 0} | \mathbb{1}_{n-1}^\top \sigma < \gamma\}$, by using $\sigma_n = \gamma - \sum_{j=1}^{n-1} \sigma_j$ and then apply the theorem to conclude existence of a differentiable point in $\{\sigma \in \mathbb{R}^{n-1}_{\geq 0} | \mathbb{1}_{n-1}^\top \sigma < \gamma\}$ which implies existence of a differentiable point in the original $\mathcal{S}_{>0}$.

$\mathcal{A}(\ell_2, \sigma^\star)\backslash\{\ell_2\}$. Note that it cannot be $\ell_3 = \ell_1$, because this means

$$u_{\ell_2}(\sigma^\star_{\ell_2}) \leq u_{\ell_1}(\sigma^\star_{\ell_1}) - c_{\ell_2\ell_1},$$

which together with (7.17) results in

$$u_{\ell_1}(\sigma^\star_{\ell_1}) < u_{\ell_1}(\sigma^\star_{\ell_1}) - c_{\ell_2\ell_1} - c_{\ell_1\ell_2},$$

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 $\mathcal{A}(\ell_3, \sigma^\star) = \{\ell_3\}$ then the statement is proven with $h^\star = \ell_2, j^\star = \ell_3$, otherwise there exists $\ell_4 \notin \{\ell_1, \ell_2, \ell_3\}$ such that $\ell_4 \in \mathcal{A}(\ell_3, \sigma^\star)$. 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 $\ell_k \in \mathcal{A}(\ell_{k-1}, \sigma^\star)$ and $\mathcal{A}(\ell_k, \sigma^\star) = \{\ell_k\}$, thus proving the statement with $h^\star = \ell_{k-1}$ and $j^\star = \ell_k$. We now show that not only $j^\star \in \mathcal{A}(h^\star, \sigma^\star)$, but actually $\mathcal{A}(h^\star, \sigma^\star) = \{j^\star\}$. For the sake of contradiction, assume that there exists $\ell \neq j^\star$ such that $\ell \in \mathcal{A}(h^\star, \sigma^\star)$. This means that $F_{h^\star}(\sigma^\star) = u_{j^\star}(\sigma^\star_{j^\star}) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star j^\star} = u_\ell(\sigma^\star_\ell) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star\ell}$. Then consider the vector of the canonical basis $\mathbf{e}_{j^\star} \in \mathbb{R}^n$ and compute

$$\begin{aligned} &\lim_{t\to 0^+} \frac{F_{h^\star}(\sigma^\star + t\mathbf{e}_{j^\star}) - F_{h^\star}(\sigma^\star)}{t} = \\ &\lim_{t\to 0^+} \frac{[u_\ell(\sigma^\star_\ell) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star\ell}] - [u_\ell(\sigma^\star_\ell) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star\ell}]}{t} = 0, \end{aligned} \tag{7.18}$$

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

$$u_{j^\star}(\sigma^\star_{j^\star} + t) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star j^\star} < u_{j^\star}(\sigma^\star_{j^\star}) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star j^\star} = u_\ell(\sigma^\star_\ell) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star\ell},$$

due to $u_{j^\star}$ being strictly decreasing by assumption. Moreover,

$$\begin{aligned} &\lim_{t\to 0^-} \frac{F_{h^\star}(\sigma^\star + t\mathbf{e}_{j^\star}) - F_{h^\star}(\sigma^\star)}{t} = \\ &\lim_{t\to 0^-} \frac{[u_{j^\star}(\sigma^\star_{j^\star} + t) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star j^\star}] - [u_{j^\star}(\sigma^\star_{j^\star}) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star j^\star}]}{t} = \\ &\lim_{t\to 0^-} \frac{u_{j^\star}(\sigma^\star_{j^\star} + t) - u_{j^\star}(\sigma^\star_{j^\star})}{t} = u'_{j^\star}(\sigma^\star_{j^\star}) < 0. \end{aligned} \tag{7.19}$$

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

$$u_{j^\star}(\sigma^\star_{j^\star} + t) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star j^\star} > u_{j^\star}(\sigma^\star_{j^\star}) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star j^\star} = u_\ell(\sigma^\star_\ell) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star\ell},$$

due to $u_{j^\star}$ being strictly decreasing by assumption. From (7.18) and (7.19) we obtain that $F_{h^\star}$ is not differentiable[2] at $\sigma^\star$, against what proved in the second part. Hence we must conclude that there cannot exist $\ell \neq j^\star$ such that $\ell \in \mathcal{A}(h^\star, \sigma^\star)$, thus $\mathcal{A}(h^\star, \sigma^\star) = \{j^\star\}$.

[2]Again, to be more rigorous one should define F in the $n-1$ dimensional domain.

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

$$\begin{aligned} u_{j^\star}(\sigma^\star_{j^\star}) - c_{h^\star j^\star} > u_\ell(\sigma^\star_\ell) - c_{h^\star \ell}, \quad \forall\, \ell \neq j^\star, \\ u_{j^\star}(\sigma^\star_{j^\star}) - c_{j^\star j^\star} > u_\ell(\sigma^\star_\ell) - c_{j^\star \ell}, \quad \forall\, \ell \neq j^\star. \end{aligned} \tag{7.20}$$

As a consequence of (7.20) there exists a small enough open ball around $\sigma^\star$ where $F_{j^\star}(\sigma^\star) = u_{j^\star}(\sigma^\star_{j^\star}) - u_{j^\star}(\sigma^\star_{j^\star}) - c_{j^\star j^\star} = 0$ and $F_{h^\star}(\sigma^\star) = u_{j^\star}(\sigma^\star_{j^\star}) - u_{h^\star}(\sigma^\star_{h^\star}) - c_{h^\star j^\star}$. Thus

$$[\nabla_\sigma F(\sigma^\star)]_{j^\star h^\star \times j^\star h^\star} = \begin{bmatrix} \frac{\partial F_{j^\star}(\sigma^\star)}{\partial \sigma_{j^\star}} & \frac{\partial F_{j^\star}(\sigma^\star)}{\partial \sigma_{h^\star}} \\ \frac{\partial F_{h^\star}(\sigma^\star)}{\partial \sigma_{j^\star}} & \frac{\partial F_{h^\star}(\sigma^\star)}{\partial \sigma_{h^\star}} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ u'_{j^\star}(\sigma^\star_{j^\star}) & -u'_{h^\star}(\sigma^\star_{h^\star}) \end{bmatrix},$$

whose symmetric part has determinant $0 \cdot (-u'_{h^\star}(\sigma^\star_{h^\star})) - (u'_{j^\star}(\sigma^\star_{j^\star}))^2/4 < 0$, which makes $[\nabla_\sigma F(\sigma^\star)]_{j^\star h^\star \times j^\star h^\star}$ indefinite. Thus $\nabla_\sigma F(\sigma^\star)$ itself is indefinite and F is not monotone in $\mathcal{S}$ due to Proposition 19. □

Remark 6. Even though the statement of Theorem 10 requires u_j to be strictly decreasing for all j, its proof only requires $u_{j^\star}$ to be strictly decreasing at $\sigma^\star_{j^\star}$. Indeed, it is enough to assume that u_j is Lipschitz for all j, $c_{jh} \geq 0$ for all $j, h \in \{1, \dots, n\}$ and there exist $\sigma^\star \in \mathcal{S}_{>0}$, $j^\star, h^\star \in \{1, \dots, n\}$ such that $j^\star \in \mathcal{A}(h^\star, \sigma^\star), \mathcal{A}(j^\star, \sigma^\star) = \{j^\star\}$ and $u_{j^\star}$ is strictly decreasing at $\sigma^\star_{j^\star}$. □

Remark 7. Definition 11 of parallel Wardrop coincides with Definition 12 of inertial Wardrop if $c_{jh} = 0$ for all $j, h \in \{1, \dots, n\}$ and, as a consequence of Lemma 14 and Theorem 9, the solution sets of VI$(\mathcal{S}, -u)$ and VI$(\mathcal{S}, F)$ coincide. This does not mean that, if $c_{jh} = 0$ for all j, h, the operators $-u$ and F exhibit the same monotonicity properties outside the solution set. Indeed, if u_j is non-increasing for all j, by the proof of Lemma 16 $-u$ is monotone, whereas F can lack monotonicity, as just stated in Theorem 10. □

We conclude this section by pointing out that both Example 4 and Example 5 satisfy the sufficient condition of Theorem 10. This reflects for Example 4 in the indefinite matrix

$$\nabla_\sigma F([0.2; 0.2; 0.6]) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

and for Example 5 in in the indefinite matrix

$$\nabla_\sigma F([0.25; 0.5; 0.25]) = \begin{bmatrix} 0 & 0 & 0 \\ -1.75 & 4.1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

7.4 Dynamics converging to an inertial Wardrop equilibrium

At the end of Section 7.2 we commented on two important drawbacks of Algorithm 6. This section is dedicated to the design of an algorithm that overcomes them. To this end, let us introduce some notation.

Given $\sigma \in \mathcal{S}$, for each j such that $\sigma_j > 0$ we define the envy set of j as

$$\mathcal{E}_j^{\text{out}}(\sigma) \coloneqq \{h \in \{1, \ldots, n\} \text{ s.t. } u_j(\sigma_j) < u_h(x_h) - c_{jh}\}.$$

If instead $\sigma_j = 0$ we define $\mathcal{E}_j^{\text{out}}(\sigma) = \emptyset$. Note that the definition of $\mathcal{E}_j^{\text{out}}$ is different from that of $\mathcal{A}(j,\sigma)$ given in (7.16). For any $j \in \{1, \ldots, n\}$, we define $\mathcal{E}_j^{\text{in}}(\sigma) = \{h \in \{1, \ldots, n\} \text{ s.t. } j \in \mathcal{E}_h^{\text{out}}(\sigma)\}$. Using this notation

$$\begin{aligned}&\bar\sigma \text{ is an inertial Wardrop equilibrium } \Leftrightarrow\\ &\bar\sigma \in \mathcal{S} \text{ and } \mathcal{E}_j^{\text{out}}(\bar\sigma) = \emptyset,\ \forall\, j \in \{1, \ldots, n\}.\end{aligned}$$

Definition 13 (Transition matrix)**.** Given $\sigma \in \mathcal{S}$ and $0 < \tau < 1$, we define the transition matrix $P(\sigma) \in \mathbb{R}^{n\times n}$ as

$$P_{jj}(\sigma) = \begin{cases} 1 & \text{if } \mathcal{E}_j^{\text{out}}(\sigma) = \emptyset \\ 1-\tau & \text{if } \mathcal{E}_j^{\text{out}}(\sigma) \neq \emptyset, \end{cases} \qquad P_{hj}(\sigma) = \begin{cases} 0 & \text{if } h \notin \mathcal{E}_j^{\text{out}}(\sigma) \\ \tau/|\mathcal{E}_j^{\text{out}}(\sigma)| & \text{if } h \in \mathcal{E}_j^{\text{out}}(\sigma), \end{cases}$$

where $|\mathcal{E}_j^{\text{out}}(\sigma)|$ indicates the cardinality of $\mathcal{E}_j^{\text{out}}(\sigma)$. □

Note that for all $\sigma \in \mathcal{S}$ we have $P(\sigma) \in \mathbb{R}^{n\times n}_{\geq 0}$ and $\mathbb{1}_n^\top P(\sigma) = \mathbb{1}_n^\top$ by construction. In other words, P is a column stochastic matrix [OSAFM07]. We are now ready to introduce the algorithm.

Algorithm 7: Improvement dynamics

Initialization:	$0 < \tau < 1$, $k = 0$, $\sigma(0) \in \mathcal{S}$.
Iterate:	$\sigma(k+1) = P(\sigma(k))\,\sigma(k)$, $k \leftarrow k+1$.

Algorithm 7 can be given the following interpretation: at iteration k, if the agents at action h see one alternative attractive action, a fraction τ of them switches to that action. If there is more than one attractive action, then the fraction of τ agents equally splits among the attractive actions.

Algorithm 7 does not present the two drawbacks of Algorithm 6 explained at the end of Section 7.2. First of all, agents switch action only if the switch is convenient. Secondly, to determine whether a switch is convenient they need to be aware of the inertial coefficients and they only need measure the other actions' utilities $u(\sigma(k))$; contrary to Algorithm 6, indeed, they do not need to measure the mass of agents $\sigma(k)$ performing other actions. As a consequence, Algorithm 7 can be interpreted as the natural dynamics of agents switching actions until they reach an equilibrium.

Since $\sigma(0) \in \mathcal{S}$, then $\sigma(k) \in \mathcal{S}$ for all k. Indeed, non-negativity is preserved because $P(\sigma) \in \mathbb{R}^{n\times n}_{\geq 0}$, and $\mathbb{1}_n^\top \sigma(k+1) = \mathbb{1}_n^\top P(\sigma(k))\sigma(k) = \mathbb{1}_n^\top \sigma(k) = \mathbb{1}_n^\top \sigma(0)$, due to column-stochasticity of $P(\sigma)$.

Theorem 11. Assume that in $[0, \gamma]$ the utility u_j is non-increasing and L-Lipschitz for each $j \in \{1, \dots, n\}$ and there exists $c_{\min} > 0$ such that $c_{jh} \geq c_{\min}$ for all $j, h \in \{1, \dots, n\}$. If $0 < \tau < c_{\min}/(L\gamma)$, then $\sigma(k)$ in Algorithm 7 converges to an inertial Wardrop equilibrium $\bar{\sigma}$. Moreover, if $\bar{\sigma} > \mathbb{0}_n$, then the algorithm terminates in a finite number of steps.

Proof. We prove the statement by showing that $\sigma(k) \to \bar{\sigma}$ such that $\mathcal{E}_j^{\text{out}}(\bar{\sigma}) = \emptyset$ for all $j \in \{1, \dots, n\}$. Let us denote $\mu(k) = \min_{j\in\{1,\dots,n\}} u_j(\sigma_j(k))$. We show in the following that $\mu(k)$ is a non-decreasing sequence.

First, note that for any action j we have $\sigma_j(k+1) - \sigma_j(k) \leq \tau\gamma$, because any other action h transfers at most $\tau\sigma_h(k)$ to action h. Then we can bound the maximum utility decrease

$$\begin{aligned} u_j(\sigma_j(k+1)) - u_j(\sigma_j(k)) &\geq -L|\sigma_j(k+1) - \sigma_j(k)| \\ &\geq -L\tau\gamma \geq -L\tau\gamma = -\beta c_{\min}, \end{aligned} \tag{7.21}$$

where the first inequality follows by Lipschitz continuity, the last by the choice of τ and $\beta = (\tau L\gamma)/c_{\min} \in]0, 1[$.

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

$$j \text{ faces utility decrease at step } k \Rightarrow \quad u_j(\sigma_j(k)) > u_h(\sigma_h(k)) + c_{hj} \geq \mu(k) + c_{\min}. \tag{7.22}$$

Combining (7.21) with (7.22) we obtain

$$\begin{gathered} j \text{ faces utility decrease at step } k \Rightarrow \\ u_j(\sigma_j(k+1)) > \mu(k) + (1-\beta)c_{\min}, \end{gathered}$$

which implies

$$\mu(k+1) \geq \mu(k).$$

Since $\mu(k)$ is non-decreasing and it has an upper bound ($\{u_j\}_{j=1}^n$ are continuous functions in a compact set), there exists a value $\mu^\star$ such that

$$\lim_{k\to\infty} \mu(k) = \mu^\star. \tag{7.23}$$

We show in the following that there exists an action $j^\star$ such that

$$\lim_{k\to\infty} u_{j^\star}(\sigma_{j^\star}(k)) = \mu^\star. \tag{7.24}$$

To this end, we note that by definition of $\lim_{k\to\infty}\mu(k)=\mu^\star$, there exists $\hat{k}$ such that

$$\mu(k) > \mu^\star - c_{\min}(1-\beta)/2, \quad \forall k \geq \hat{k}. \tag{7.25}$$

Then

$$\begin{aligned} &j \text{ faces utility decrease at step } k \geq \hat{k} \Rightarrow \\ &u_j(\sigma_j(k)) \geq \mu^\star - c_{\min}(1-\beta)/2 + c_{\min} \\ &\;= \mu^\star + c_{\min}(1+\beta)/2, \end{aligned} \tag{7.26}$$

where the first inequality follows from combining (7.22) and (7.25). Combining (7.21) and (7.26) we obtain

$$\begin{aligned} &j \text{ faces utility decrease at step } k \geq \hat{k} \Rightarrow \\ &u_j(\sigma_j(k+1)) \geq \mu^\star - c_{\min}(1-\beta)/2 + c_{\min}(1-\beta) \\ &\;= \mu^\star + c_{\min}(1-\beta)/2. \end{aligned} \tag{7.27}$$

Figure 7.3 illustrates the inequalities (7.26) and (7.27).

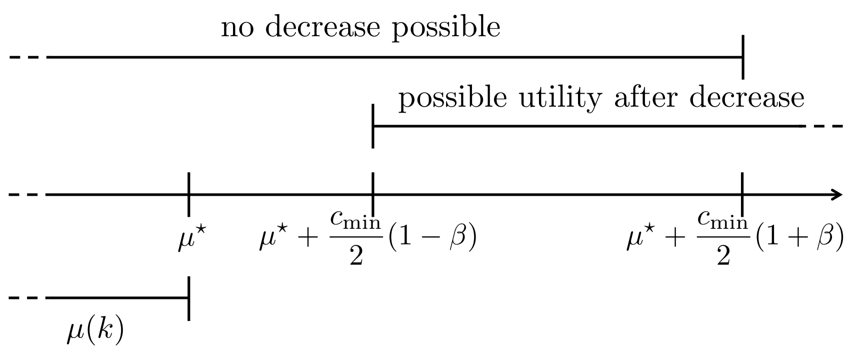


Figure 7.3: Illustration of $\mu(k) \to \mu^\star$ from below and of the inequalities (7.26) and (7.27) after iteration $\hat{k}$ (with $\beta = 0.5$).

Combining inequalities (7.26) and (7.27) we obtain that

$$\begin{aligned} &\exists\, k_1 \geq \hat{k} \text{ such that } u_j(\sigma_j(k_1)) \geq \mu^\star + \rho > \mu^\star \Rightarrow \\ &u_j(\sigma_j(k)) \geq \min\{\mu^\star + \rho, \mu^\star + c_{\min}(1-\beta)/2\} \text{ for all } k \geq k_1. \end{aligned} \tag{7.28}$$

It then follows

$$\exists\, k_1 \geq \hat{k} \text{ such that } u_j(\sigma_j(k_1)) > \mu^\star \Rightarrow \lim_{k\to\infty} u_j(\sigma_j(k)) \neq \mu^\star. \tag{7.29}$$

By (7.29) and (7.23) it follows that there exists at least an action $j^\star$ such that it holds $u_{j^\star}(\sigma_{j^\star}(k)) \leq \mu^\star$ for all $k \geq \hat{k}$. Using again (7.23) and the squeeze theorem, we can conclude that $j^\star$ satisfies (7.24).

Note that for $k \geq \hat{k}$ the set $\mathcal{E}^{\text{in}}_{j^\star}(\sigma(k))$ is empty due to (7.22) and $u_{j^\star}(\sigma_{j^\star}) \leq \mu^\star$. In words, no other action can envy $j^\star$ after step $\hat{k}$. This implies that $u_{j^\star}(\sigma_{j^\star}(k))$ is a non-decreasing sequence, and in turn $\sigma_{j^\star}(k)$ is a non-increasing sequence, due to the fact that $u_{j^\star}$ is a non-increasing function. As a consequence

$$\lim_{k\to\infty} \sigma_{j^\star}(k) = \bar{\sigma}_{j^\star} \geq 0. \tag{7.30}$$

If $\bar{\sigma}_{j^\star} = 0$, then clearly $\mathcal{E}^{\text{out}}_{j^\star}(\bar{\sigma}_{j^\star}, \sigma_{-j^\star}) = \emptyset$ by definition, independently from the values taken by $\sigma_{-j^\star}$. If instead $\bar{\sigma}_{j^\star} > 0$, since the only possible mass decrease for $j^\star$ is of the form $\sigma_{j^\star}(k+1) = (1-\tau)\sigma_{j^\star}(k)$, then convergence is achieved in a finite number of steps. In other words, there exists $\tilde{k}$ such that $\sigma_{j^\star}(k) = \bar{\sigma}_{j^\star}$ for all $k \geq \tilde{k}$. In this case, for $k \geq \tilde{k}$ not only $\mathcal{E}^{\text{in}}_{j^\star}(\sigma(k)) = \emptyset$, but also $\mathcal{E}^{\text{out}}_{j^\star}(\sigma(k)) = \emptyset$, because otherwise $j^\star$ would encounter a mass decrease.

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

The last argument distinguishes two cases: $\bar{\sigma}_{j^\star} > 0$ and $\bar{\sigma}_{j^\star} = 0$. In the first case $\bar{\sigma}_{j^\star} > 0$, we already showed that there exists $\tilde{k}$ such that $\mathcal{E}^{\text{in}}_{j^\star}(\sigma(k)) = \mathcal{E}^{\text{out}}_{j^\star}(\sigma(k)) = \emptyset$ for all $k > \tilde{k}$. Then action $j^\star$ has no interaction with any the other action and considering $k \geq \tilde{k}$ we apply to $\{1, \ldots, n\} \backslash j^\star$ the reasoning of part (ii) of this proof until equation (7.30) to show that there is an action $h^\star \in \{1, \ldots, n\} \backslash \{j^\star\}$ with mass that converges to $\bar{\sigma}_{h^\star}$ (in a finite number of steps if $\bar{\sigma}_{h^\star} > 0$).

Let us now focus on the second case $\bar{\sigma}_{j^\star} = 0$. The main idea here is that, even though $\mathcal{E}^{\text{out}}_{j^\star}$ does not become the empty set at any finite iteration k, the mass $\sigma_{j^\star}$ 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 $\eta(k) = \min_{h \in \{1,\ldots,n\} \backslash \{j^\star\}} u_h(\sigma_h(k))$. Contrary to $\mu(k)$, the sequence $\eta(k)$ is not non-decreasing in general because the analogous of (7.22) does not hold, as action $j^\star$ could transfer some of its mass to $\{1, \ldots, n\} \backslash \{j^\star\}$ thus making their utilities decrease. Nonetheless, we show that there exists $\eta^\star$ such that

$$\lim_{k\to\infty} \eta(k) = \eta^\star. \tag{7.31}$$

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

$$\sigma_{j^\star}(k) < \varepsilon/(2L), \quad \forall\, k \geq k_\infty. \tag{7.32}$$

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 $j^\star$, 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 $\eta^\star$. By definition, there exists k^0 such that $\eta^0(k) > \eta^\star - \varepsilon/2$ for all $k \geq k^0$. Moreover, $\delta(k+1) - \delta(k) = \max\{0, \eta(k) - \eta(k+1)\} > 0$ only if $\mathcal{E}^{\text{out}}_{j^\star}(\sigma(k)) \neq \emptyset$ and in this case $\max\{0, \eta(k) - \eta(k+1)\} \leq L \cdot \tau\sigma_j(k)$. In words, the only way $\eta(k)$ can decrease is if action $j^\star$ transfers some mass to the others, and even then we have a bound on the utility decrease that this can cause. Summing up

$$\lim_{k\to\infty} \delta(k) = \sum_{k=k_\infty}^{\infty} \max\{0, \eta(k) - \eta(k+1)\} \leq L\sigma_{j^\star}(k_\infty) \underset{(7.32)}{<} \varepsilon/2,$$

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

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

which proves (7.31).

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

$$\lim_{k\to\infty} u_{h^\star}(\sigma_{h^\star}(k)) = \eta^\star. \tag{7.33}$$

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

$$\lim_{k\to\infty} u_\ell(\sigma_\ell(k)) \neq \eta^\star. \tag{7.34}$$

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 (7.34) is equivalent to the existence of $\theta > 0$ such that for all $k' \geq 0$ there exists $k'' \geq k'$ such that

$$u_\ell(\sigma_\ell(k'')) > \eta^\star + \theta. \tag{7.35}$$

There are two possibilities in which ℓ can face a utility decrease after k'', namely through a mass transfer from some action $\{1, \dots, n\} \backslash \{j^\star, \ell\}$ or through a mass transfer from action $j^\star$. If the mass transfer happens through some action $\{1, \dots, n\} \backslash \{j^\star, \ell\}$, we can use the same argument of Figure 7.3 and in particular of implication (7.28) to conclude from (7.35) that

$$u_\ell(\sigma_\ell(k)) \geq \min\{\eta^\star + \theta, \eta^\star + c_{\min}(1-\beta)/2\}, \ \forall\, k \geq k''. \tag{7.36}$$

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

$$\sigma_{j^\star}(k) < \theta/(2L), \quad \forall k \geq k' \tag{7.37}$$

and take k'' such that (7.35) holds. Then

$$u_\ell(\sigma_\ell(k)) \geq u_\ell(\sigma_\ell(k'')) - L\frac{\theta}{2L} > \eta^\star + \theta - \frac{\theta}{2} = \eta^\star + \frac{\theta}{2}. \tag{7.38}$$

for all $k \geq k''$, where the first inequality holds due to Lipschitz continuity and to (7.37), while the second inequality holds due to (7.35). We can conclude that if (7.34) holds for action ℓ, then either (7.36) or (7.38) holds. Consequently, after k'' action ℓ does not attain the minimum $\eta(k)$. If (7.34) holds for all $\ell \in \{1,\dots,n\}\backslash j^\star$, then the minimum $\eta(k)$ is not attained by any action after k'', which is a contradiction. Then there must exist $h^\star$ such that (7.33) holds. With the same argument that led to (7.30), we can conclude that there exists $\bar{\sigma}_{h^\star} \geq 0$ such that $\lim_{k\to\infty} \sigma_{h^\star}(k) = \bar{\sigma}_{h^\star} \geq 0$. As done for $j^\star$, we can conclude that $\mathcal{E}^{\text{out}}_{h^\star} = \emptyset$. □

Generalizations

We present in the following three generalizations of the material presented in this chapter.

Multi-class inertial Wardrop equilibrium

The concept of inertial Wardrop introduced in Definition 12 relies on the idea that each infinitesimal agent perceives the same utility u_j and the same inertial coefficients c_{jh}. This assumption can be relaxed by introducing A different classes and indicating with σ_j^α the mass of agents belonging to class α which choose action j. We denote $\sigma_j = \sum_{\alpha=1}^A \sigma_j^\alpha$ and $\sigma^\alpha = \{\sigma_j^\alpha\}_{j=1}^n$.

Definition 14. Consider utilities $\{u_j^\alpha : \mathbb{R} \to \mathbb{R}\}$, inertia coefficients $\{c_{jh}^\alpha \geq 0\}$ and masses $\{\gamma^\alpha > 0\}$, with $j, h \in \{1,\dots,n\}$, $\alpha \in \{1,\dots,A\}$. The vector $\bar{\sigma} \in \mathbb{R}^{nA}$ is a *multi-class inertial Wardrop equilibrium* if $\bar{\sigma} \geq \mathbb{0}_{nA}$, $\mathbb{1}_A^\top \bar{\sigma}^\alpha = \gamma^\alpha$ for all α, and

$$\bar{\sigma}_j^\alpha > 0 \Rightarrow u_j^\alpha(\bar{\sigma}_j) \geq u_h^\alpha(\bar{\sigma}_h) - c_{jh}^\alpha, \quad \forall\, h \in \{1,\dots,n\}, \tag{7.39}$$

for all $j \in \{1,\dots,n\}$ and $\alpha \in \{1,\dots,A\}$. □

Note that even though different classes might perceive different utilities at the same action h, each of these utilities is a function of the sole $\sum_{\alpha=1}^A \sigma_j^\alpha$. This is indeed what couples the different classes together. If in Definition 14 we have $c_{jh} = 0$ for all j, h, we obtain the definition of *multi-class parallel Wardrop equilibrium*. Upon redefining

$\mathcal{S} = \mathcal{S}^1 \times \cdots \times \mathcal{S}^A \subset \mathbb{R}^{nA}$ as the Cartesian product of the simplex sets, one can redefine the operator $F : \mathcal{S} \to \mathbb{R}^{nA}_{\geq 0}$, where

$$F(\sigma) = [[F^\alpha_j(\sigma)]^A_{\alpha=1}]^n_{j=1},$$
$$F^\alpha_j(\sigma) = \max_{h\in\{1,\dots,n\}} (u^\alpha_h(\sigma_h) - u^\alpha_j(\sigma_j) - c^\alpha_{jh}).$$

Using a trivial extension of the proof of Theorem 9, it is possible to show that the set of multi-class parallel Wardrop equilibria coincides with the solution set of VI(S, F). Theorem 10 about lack of monotonicity also extends to the multi-class case. The first two parts of the proof are identical, the last two parts only differ in that $j^\star$, $h^\star$ exist for a specific class $\alpha^\star$.

Algorithm 7 can also be extended. Indeed, let us define

$$\mathcal{E}^{\text{out},\alpha}_j(\sigma) \coloneqq \left\{h \in \{1,\dots,n\} \text{ s.t. } u^\alpha_j(\sigma_j) < u^\alpha_h(x_h) - c^\alpha_{jh}\right\}$$

and $P \in \mathbb{R}^{nA\times nA} = \text{blkdiag}(P^1,\dots,P^A)$, where the entries of P^α are

$$P^\alpha_{jj}(\sigma) = \begin{cases} 1 & \text{if } \mathcal{E}^{\text{out},\alpha}_j(\sigma) = \emptyset \\ 1-\tau & \text{if } \mathcal{E}^{\text{out},\alpha}_j(\sigma) \neq \emptyset, \end{cases} \qquad P^\alpha_{hj}(\sigma) = \begin{cases} 0 & \text{if } h \notin \mathcal{E}^{\text{out},\alpha}_j(\sigma) \\ \tau/|\mathcal{E}^{\text{out},\alpha}_j(\sigma)| & \text{if } h \in \mathcal{E}^{\text{out},\alpha}_j(\sigma). \end{cases}$$

Then $\sigma(k+1) = P(\sigma(k))\sigma(k)$ converges to a multi-class parallel Wardrop equilibrium, under the same assumptions of Theorem 11. Some parts of the proof have to be slightly generalized but the structure remains the same. We do not report the extension of the proof in full detail.

Relaxations of Algorithm 7

When introducing Algorithm 7 we specified that, in presence of multiple alternative attractive actions, the fraction τ of agents equally splits among the attractive actions. This restriction is not needed, i.e., the fraction τ can split in any arbitrary way among the attractive actions without compromising the convergence result.

Moreover, the proof never uses the fact that the agents switch in a synchronous manner. As a consequence, Algorithm 7 converges also if the agents switch asynchronously.

Finally, instead of assuming that a τ fraction of the agents switches and requiring $\tau < c_{\min}/L\gamma$, one could let any number of agents switch as long as this number is bounded by $\tau < c_{\min}/L$.

Atomic agents with discrete action set

Instead of a continuum of infinitesimal agents, one could consider a finite number M of atomic agents. Each agent possesses unitary mass and can choose only one of the actions $\{1,\dots,n\}$. In other words, his strategy x^i must belong to the set $\{\mathbf{e}_j\}^n_{j=1}$, where $\mathbf{e}_j$ is the j^{th} vector of the canonical basis. The utility u_j is a function of $\sigma_j = (1/M)\sum_{i=1}^{M} x_j^i$. The definition of inertial equilibrium then requires that no atomic agent $i \in \{1, \ldots, M\}$ has an incentive to switch action, considering the utilities of the alternative actions and the corresponding inertial coefficients. The continuum of infinitesimal agents studied in this chapter represents the limiting situation obtained as the number of agents M grows to infinity. Note that the theory of Chapters 2-3 does not apply, as the action space is discrete. As a consequence, the reformulation as VI is not possible. Nonetheless, one can formulate Algorithm 7 by letting an agent i switch to an arbitrary action whenever such action is attractive. Convergence is guaranteed upon substituting the original bound on τ with a bound on the maximum number of agents that can switch at the same time.

Comparison with the literature

To the best of our knowledge, the concept of inertial Wardrop equilibrium in Definition 12 is novel, due to the presence of the inertial coefficients c_{jh}. As a consequence, all the material in the chapter has not been previously studied by other authors.

The journey that leads us to the definition of the inertial Wardrop equilibrium starts with the study of the *migration equilibrium* introduced by Anna Nagurney. The series of works [Nag89, Nag89, NPZ92, NPZ93] defines the migration equilibrium in a way that resembles Definition 12, but with some important differences. The migration equilibrium considers a fixed initial mass distribution $\sigma^0 \in \mathcal{S}$, with σ_j^0 representing the amount of agents residing at a physical location j, which enjoy utility $u_j(\sigma^0)$. The initial distribution σ^0 is transformed into the final distribution $\sigma^1 \in \mathcal{S}$, which is a function of the migrations $(f_{jh})_{j,h=1}^n$ (that in our terms are action switches). The variables are thus the migrations $(f_{jh})_{j,h=1}^n$ themselves. Each migration comes with a migration cost $c_{jh}(f_{jh})$ which is a function of the amount f_{jh} of agents migrating. In our case instead c_{jh} is constant. A migration equilibrium consists in a set of migrations $(f_{jh})_{j,h=1}^n$ such that, considering the fixed initial utilities $u(s^0)$, the migration costs $c_{jh}(f_{jh})$ and the corresponding final utilities $u(\sigma^1)$, no other set of migrations is more convenient.

To sum up, the migration equilibrium and the inertial Wardrop of Definition 12 share the same key ingredients, which are however arranged in a different manner. The main consequences are that the migration equilibrium equivalent VI is different from our VI in Theorem 9 and, most importantly, the algorithms proposed in the works of Nagurney do not resemble Algorithm 7. In particular, they do not aim at describing the natural dynamics of the agents seeking an equilibrium.

On the other hand, the works in *population games* focus exactly on (continuous-time) agents dynamics that achieve the Wardrop equilibrium[3] described in conditions (3.10). A particular class of agents dynamics is those of imitation dynamics, which are similar to

[3] In the population games literature a point satisfying conditions (3.10) is referred to as Nash equilibrium [San10, p. (24)], contrary to the terminology used in this thesis.

the discrete-time Algorithm 7 in that agents switch to more attractive actions. Different works provide local [San01, Nac90] and global [CT14, ZCF17] convergence guarantees, but rather than digging into the vast literature of population games, we point out that those works consider inertia coefficients equal to zero, hence study a different problem. As highlighted in Chapter 8, adapting the agents dynamics proposed in population games to the case of positive inertia coefficients is a very interesting possibility.

Finally, we remark that introducing different classes of agents is not a novel idea, because it appears already in the migration equilibrium works [Nag13, Section 5.3] and in the population games ones [San10, Section 2.1].

CHAPTER 8

Conclusions and outlook

This thesis focused on various aspects of equilibrium problems in aggregative games. Chapter 1 motivated the relevance of the systems that are described by aggregative games. Chapter 2 presented background material, while Chapter 3 was dedicated to the systematic study of the relations between Nash and Wardrop equilibria through the framework of variational inequalities. Chapters 4 and 5 designed parallel and distributed algorithms that converge to such equilibria. Chapter 6 focused in detail on three different case studies to verify the results of the previous Chapters 3-5. Chapter 7 introduced the novel concept of Wardrop equilibrium with inertia and proposed agent dynamics that converge to it.

In the following we make clear what we believe are the main novel technical contributions of the thesis by listing them one by one.

- Theorem 1 provided sufficient conditions for strong monotonicity of the Nash operator for a diagonal price function;

- Theorem 2 bounded the distance between the strategies at the Nash and the strategies at the Wardrop in terms of the number of agents;

- Theorem 3 and Theorem 4 proved convergence in presence of coupling constraints of the parallel best-response Algorithm 3 and of the parallel gradient-step Algorithm 4, respectively;

- Theorem 5 showed that the distributed gradient-step Algorithm 5 converges to an almost-Nash equilibrium in presence of coupling constraints;

- Theorem 10 provided weak sufficient conditions for absence of monotonicity of the VI relative to the inertial Wardrop equilibrium;

- Theorem 11 established convergence to an inertial Wardrop equilibrium of the agents dynamics in Algorithm 7.

Outlook

Regarding future research directions, for almost each one of the sections of the thesis we could propose some plausible further work. We list in the following the three future developments that we consider the most valuable and on which we would invest our time.

Distributed algorithm for exact Nash

Chapter 5 proposed the distributed Algorithm 5 whose iteration requires ν communications among the agents. Theorem 5 proved that, for each $\varepsilon > 0$, there exists $\nu_\varepsilon > 0$ such that by setting $\nu > \nu_\varepsilon$, Algorithm 5 converges to an ε-Nash equilibrium. Figure 6.12 suggests that in practical situations the number ν of communications required to achieve a good accuracy ε might not be very high. Nonetheless, from a theoretical point of view it would be valuable to design an algorithm that achieves an exact Nash equilibrium (rather than an ε-Nash) with only one communication (i.e., $\nu = 1$) per iteration, over a generic strongly connected and doubly-stochastic communication network. In particular, one would need to propose a gradient step algorithm and a best-response algorithm. Designing the two algorithms would probably require two different approaches, as indeed was the case for the parallel Algorithms 3 and 4.

Enhancing the results of the inertial Wardrop equilibrium

One open question about the inertial Wardrop equilibrium of Chapter 7 is whether the equilibrium set is connected or not. The three-dimensional examples that we investigated so far visually indicate that the answer is positive, as in Figure 7.1 and Figure 7.2. When we instead simulated action spaces of considerably higher cardinality, it was difficult to establish connectedness of the solution set, due to the lack of a visual representation similar to those of Figure 7.1 and Figure 7.2.

It would be also important to make the generalizations presented at the end of Section 7.4 into rigorous statements. In particular, we believe that the proof of convergence of Algorithm 7 might simplify a bit if, instead of assuming that a τ fraction of the agents switches and requiring $\tau < c_{\min}/L\gamma$, one could let any number of agents switch as long as this number is bounded by $\tau < c_{\min}/L$. In particular, one could probably prove with a very simple example that such bound is tight, in the sense that there exists an instance of utilities and positive inertial coefficients such that the modified version of Algorithm 7 oscillates whenever $\tau \geq c_{\min}/L$.

Finally, finding a relevant application of the results of Chapter 7 would be extremely valuable. In the work [GPO17] we have developed a case study on area coverage for taxi drivers in the territory of Hong Kong, but the model does not necessarily describe the taxi drivers' behavior in a realistic manner. The concept of migration equilibrium of Anna Nagurney has been applied in ecology to study migration patterns of tuna fish [MKM16]. We believe that the Wardrop equilibrium with inertia could find application in planning of technological systems used by a large number of people. In the terminology of Chapter 7, modifying the perceived utilities (through economic incentives) or the inertial coefficients (through fees or tolls), or even including a new action in the game (through enhancement of the current infrastructure), would result in the old equilibrium configuration to not be an equilibrium anymore. Since the new configuration will in general possess a plethora of equilibria, the agent dynamics of Algorithm 7 could be used to predict to which specific equilibrium the agents will move to.

More general concept of equilibrium with inertia

The presence of the inertia coefficients introduced in Chapter 7 for non-atomic agents could be incorporated in the standard Definitions 9 and 10 of Nash and Wardrop equilibrium expressed for M atomic agents. A set of strategies $x_{\mathrm{N}} = [x_{\mathrm{N}}^1; \dots ; x_{\mathrm{N}}^M] \in \mathbb{R}^{Mn}$ is then an inertial Nash equilibrium if for all $i \in \{1, \dots, n\}$ and $x^i \in \mathcal{X}^i$

$$J^i(x_{\mathrm{N}}^i, \sigma(x_{\mathrm{N}})) \leq J^i\left(x^i, \tfrac{1}{M}x^i + \tfrac{1}{M}\textstyle\sum_{j\neq i} x_{\mathrm{N}}^j\right) + f^i(x_{\mathrm{N}}^i, x^i), \tag{8.1}$$

where f^i represents the burden for agent i of switching from action x_{N}^i to action x^i. In a more general version, one could even let f^i depend also on $\sigma(x)$. The analogous of condition (8.1) defines the Wardrop equilibrium with inertia for atomic agents.
A key step in the development of the results of Chapters 3-5 is Proposition 14, which allows to reformulate the Nash equilibrium as a variational inequality. Condition (8.1) cannot be directly reformulated as a VI, but perhaps one could exploit one of the many generalization of the concept of variational inequality, see [Noo98].

Bibliography

[ABEA06] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, *A survey on networking games in telecommunications*, Computers & Operations Research **33** (2006), no. 2, 286–311.

[ABS02a] T. Alpcan, T. Başar, R. Srikant, and E. Altman, *CDMA uplink power control as a noncooperative game*, Wireless Networks **8** (2002), no. 6, 659–670.

[ABS02b] E. Altman, T. Basar, and R. Srikant, *Nash equilibria for combined flow control and routing in networks: Asymptotic behavior for a large number of users*, IEEE Transactions on automatic control **47** (2002), no. 6, 917–930.

[ACA11] E. Altman, R. Combes, Z. Altman, and S. Sorin, *Routing games in the many players regime*, Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools (2011), pp. 525–527.

[AD54] K. J. Arrow and G. Debreu, *Existence of an equilibrium for a competitive economy*, Econometrica: Journal of the Econometric Society (1954), 265–290.

[ADPL94] R. Arnott, A. De Palma, and R. Lindsey, *The welfare effects of congestion tolls with heterogeneous commuters*, Journal of Transport Economics and Policy (1994), 139–161.

[AFP00] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, vol. 254, Clarendon Press Oxford, 2000.

[AS94] R. Arnott and K. Small, *The economics of traffic congestion*, American scientist **82** (1994), no. 5, 446–455.

[Aub07] J.-P. Aubin, *Mathematical methods of game and economic theory*, Courier Corporation, 2007.

[AW04] E. Altman and L. Wynter, *Equilibrium, games, and pricing in transportation and telecommunication networks*, Networks and Spatial Economics **4** (2004), no. 1, 7–21.

[BC10] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2010.

[BD95] M. J. Best and B. Ding, *On the Continuity of the Minimum in Parametric Quadratic Programs*, Journal of Optimization Theory and Applications **86** (1995), no. 1, 245–250.

[Ben74] A. Bensoussan, *Points de Nash dans le cas de fonctionnelles quadratiques et jeux differentiels lineaires a n personnes*, SIAM Journal on Control and Optimization **12** (1974), no. 3, 460.

[Ber99] D. P. Bertsekas, *Nonlinear programming*, Athena scientific Belmont, 1999.

[Ber07] V. Berinde, *Iterative approximation of fixed points*, Springer, 2007.

[BG17] G. Belgioioso and S. Grammatico, *On convexity and monotonicity in generalized aggregative games*, IFAC-PapersOnLine **50** (2017), no. 1, 14338–14343.

[BMD02] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, *The explicit linear quadratic regulator for constrained systems*, Automatica **38** (2002), no. 1, 3–20.

[BMW56] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the economics of transportation*, Tech. report, 1956.

[Boo63] J. C. Boot, *On sensitivity analysis in convex quadratic programming problems*, Operations Research **11** (1963), no. 5, 771–786.

[BP13] D. Bauso and R. Pesenti, *Mean field linear quadratic games with set up costs*, Dynamic Games and Applications **3** (2013), no. 1, 89–104.

[BPC11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends® in Machine Learning **3** (2011), no. 1, 1–122.

[BPG17] G. Burger, D. Paccagnan, B. Gentile, and J. Lygeros, *Guarantees of convergence to a dynamic user equilibrium for a single arc network*, IFAC-PapersOnLine **50** (2017), no. 1, 9674 – 9679, 20th IFAC World Congress.

[Bri02] T. Brinkhoff, *A framework for generating network-based moving objects*, GeoInformatica **6** (2002), no. 2, 153–180.

[BT97] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Athena Scientific, 1997.

[Bul18] F. Bullo, *Lectures on network systems*, Version 0.96, 2018, With contributions by J. Cortes, F. Dorfler, and S. Martinez.

[BV04] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[CK05] G. Christodoulou and E. Koutsoupias, *The price of anarchy of finite congestion games*, Proceedings of the thirty-seventh annual ACM symposium on Theory of computing (2005), ACM, pp. 67–73.

[Cla90] F. H. Clarke, *Optimization and nonsmooth analysis*, SIAM, 1990.

[CLL14] H. Chen, Y. Li, R. H. Y. Louie, and B. Vucetic, *Autonomous demand side management based on energy consumption scheduling and instantaneous load billing: An aggregative game approach*, IEEE Transactions on Smart Grid **5** (2014), no. 4, 1744–1754.

[Cot66] R. W. Cottle, *Nonlinear programs with positively bounded Jacobians*, SIAM Journal on Applied Mathematics **14** (1966), no. 1, 147–158.

[CP02] P. L. Combettes and T. Pennanen, *Generalized mann iterates for constructing fixed points in Hilbert spaces*, Journal of Mathematical Analysis and Applications **275** (2002), no. 2, 521–536.

[CSM11] J. R. Correa and N. E. Stier-Moses, *Wardrop equilibria*, Wiley encyclopedia of operations research and management science (2011).

[CSSM04] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *Selfish routing in capacitated networks*, Mathematics of Operations Research **29** (2004), no. 4, 961–976.

[CT14] R. Cressman and Y. Tao, *The replicator equation and other game dynamics*, Proceedings of the National Academy of Sciences **111** (2014), no. Supplement 3, 10810–10817.

[Daf80] S. Dafermos, *Traffic equilibrium and variational inequalities*, Transportation science **14** (1980), no. 1, 42–54.

[Daf88] ———, *Sensitivity analysis in variational inequalities*, Mathematics of Operations Research **13** (1988), no. 3, 421–434.

[DHZ06] P. Dubey, O. Haimanko, and A. Zapechelnyuk, *Strategic complements and substitutes, and potential games*, Games and Economic Behavior **54** (2006), no. 1, 77–94.

[DN87] S. Dafermos and A. Nagurney, *Oligopolistic and competitive behavior of spatially separated markets*, Regional Science and Urban Economics **17** (1987), no. 2, 245 – 254.

[DPP05] A. De Palma and N. Picard, *Route choice decision under travel time uncertainty*, Transportation Research Part A: Policy and Practice **39** (2005), no. 4, 295–324.

[DTL96] S. Dinh The Luc, Schaible, *Generalized monotone nonsmooth maps*, Journal of Convex Analysis **3** (1996), 195–206.

[FBS93] T. L. Friesz, D. Bernstein, T. E. Smith, R. L. Tobin, and B. W. Wie, *A variational inequality formulation of the dynamic network user equilibrium problem*, Operations Research **41** (1993), no. 1, 179–191.

[FFP07] F. Facchinei, A. Fischer, and V. Piccialli, *On generalized Nash games and variational inequalities*, Operations Research Letters **35** (2007), no. 2, 159–164.

[FK07] F. Facchinei and C. Kanzow, *Generalized Nash equilibrium problems*, 4OR **5** (2007), no. 3, 173–210.

[FP03] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*, Operations Research, Springer, 2003.

[GB14] P. Giselsson and S. Boyd, *Metric selection in Douglas-Rachford splitting and ADMM*, arXiv preprint arXiv:1410.8479 (2014).

[GD08] N. Geroliminis and C. F. Daganzo, *Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings*, Transportation Research Part B: Methodological **42** (2008), no. 9, 759 – 770.

[GPC16] S. Grammatico, F. Parise, M. Colombino, and J. Lygeros, *Decentralized convergence to Nash equilibria in constrained deterministic mean field control*, IEEE Transactions on Automatic Control **61** (2016), no. 11, 3315–3329.

[GPO17] B. Gentile, D. Paccagnan, B. Ogunsola, and J. Lygeros, *A novel concept of equilibrium over a network*, Proceedings of the IEEE Conference on Decision and Control (2017), IEEE, pp. 6123–6128.

[GPP17] B. Gentile*, F. Parise*, D. Paccagnan*, M. Kamgarpour, and J. Lygeros, *Nash and Wardrop equilibria in aggregative games with coupling constraints*, arXiv preprint arXiv:1702.08789 (2017).

[Gra17] S. Grammatico, *Dynamic control of agents playing aggregative games with coupling constraints*, IEEE Transactions on Automatic Control (2017).

[GTL13] L. Gan, U. Topcu, and S. H. Low, *Optimal decentralized protocol for electric vehicle charging*, IEEE Transactions on Power Systems **28** (2013), no. 2, 940–951.

[Har91] P. T. Harker, *Generalized Nash games and quasi-variational inequalities*, European journal of Operational research **54** (1991), no. 1, 81–94.

[HCM07] M. Huang, P. E. Caines, and R. P. Malhamé, *Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ε-Nash equilibria*, IEEE Transactions on Automatic Control **52** (2007), no. 9, 1560–1571.

[HM85] A. Haurie and P. Marcotte, *On the relationship between Nash-Cournot and Wardrop equilibria*, Networks **15** (1985), no. 3, 295–308.

[HRT07] A. G. Hadigheh, O. Romanko, and T. Terlaky, *Sensitivity analysis in convex quadratic optimization: Simultaneous perturbation of the objective and right-hand-side vectors*, Algorithmic Operations Research **2** (2007), no. 2, 94.

[HS66] P. Hartman and G. Stampacchia, *On some non-linear elliptic differential-functional equations*, Acta mathematica **115** (1966), no. 1, 271–310.

[Jen10] M. K. Jensen, *Aggregative games and best-reply potentials*, Economic theory **43** (2010), no. 1, 45–66.

[KCM11] T. K. Kristoffersen, K. Capion, and P. Meibom, *Optimal charging of electric drive vehicles in a market environment*, Applied Energy **88** (2011), no. 5, 1940–1948.

[KH12] S. Kar and G. Hug, *Distributed robust economic dispatch in power systems: A consensus + innovations approach*, 2012 IEEE Power and Energy Society General Meeting, pp. 1–8.

[KHR02] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, *Sumo (simulation of urban mobility)-an open-source traffic simulation*, Proceedings of the 4th Middle East Symposium on Simulation and Modelling (2002), pp. 183–187.

[KNS12] J. Koshal, A. Nedić, and U. V. Shanbhag, *A gossip algorithm for aggregative games on graphs*, Proceedings of the IEEE Conference on Decision and Control (2012), pp. 4840–4845.

[KNS16] J. Koshal, A. Nedić, and U. V. Shanbhag, *Distributed algorithms for aggregative games on graphs*, Operations Research **64** (2016), no. 3, 680–704.

[Kyp85] J. Kyparisis, *On uniqueness of Kuhn-Tucker multipliers in nonlinear programming*, Mathematical Programming **32** (1985), 242–246.

[Kyp87] J. Kyparisis, *Sensitivity analysis framework for variational inequalities*, Mathematical Programming **38** (1987), no. 2, 203–213.

[LS02] H. K. Lo and W. Szeto, *A cell-based variational inequality formulation of the dynamic user optimal assignment problem*, Transportation Research Part B: Methodological **36** (2002), no. 5, 421 – 443.

[LYH16] S. Liang, P. Yi, and Y. Hong, *Distributed Nash equilibrium seeking for aggregative games with coupled constraints*, arXiv preprint arXiv:1609.02253 (2016).

[LZ11] J. Li and H. M. Zhang, *Fundamental diagram of traffic flow : New identification scheme and further evidence from empirical data*, Journal of the Transportation Research Board (2011).

[MAS09] J. R. Marden, G. Arslan, and J. S. Shamma, *Cooperative control and potential games*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics **39** (2009), no. 6, 1393–1407.

[MCH13] 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 **21** (2013), no. 1, 67–78.

[MKM16] P. Mariani, V. Křivan, B. R. MacKenzie, and C. Mullon, *The migration game in habitat network: the case of tuna*, Theoretical Ecology **9** (2016), no. 2, 219–232.

[Mos69] U. Mosco, *Convergence of variational of convex sets and of solutions inequalities*, Advances in Mathematics **3** (1969), no. 4, 510–585.

[MS96] D. Monderer and L. S. Shapley, *Potential games*, Games and Economic Behavior **14** (1996), no. 1, 124–143.

[MW95] P. Marcotte and J. H. Wu, *On the convergence of projection methods: application to the decomposition of affine variational inequalities*, Journal of Optimization Theory and Applications **85** (1995), no. 2, 347–362.

[Nac90] J. H. Nachbar, *Evolutionary selection dynamics in games: Convergence and limit properties*, International journal of game theory **19** (1990), no. 1, 59–89.

[Nag89] A. Nagurney, *Migration equilibrium and variational inequalities*, Economics letters **31** (1989), no. 1, 109–112.

[Nag90] ———, *A network model of migration equilibrium with movement costs*, Mathematical and Computer Modelling **13** (1990), no. 5, 79–88.

[Nag13] ———, *Network economics: A variational inequality approach*, vol. 10, Springer Science & Business Media, 2013.

[Nas51] J. Nash, *Non-cooperative games*, Annals of mathematics (1951), 286–295.

[Noo98] M. A. Noor, *Generalized variational-like inequalities*, Mathematical and computer modelling **27** (1998), no. 3, 93–101.

[NPZ92] A. Nagurney, J. Pan, and L. Zhao, *Human migration networks*, European journal of operational research **59** (1992), no. 2, 262–274.

[NPZ93] A. Nagurney, J. Pan, and L. Zhao, *Human migration networks with class transformations*, Structure and Change in the Space Economy (1993), Springer, pp. 239–258.

[NRT07] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*, vol. 1, Cambridge University Press Cambridge, 2007.

[OR00] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, SIAM, 2000.

[OSAFM07] R. Olfati-Saber, J. Alex Fax, and R. M. Murray, *Consensus and cooperation in networked multi-agent systems*, Proceedings of the IEEE **95** (2007), no. 1, 215–233.

[PAE13] M. Patriksson, N. Andréasson, A. Evgrafov, E. Gustavsson, and M. Önnheim, *Introduction to continuous optimization*, Studentlitteratur, 2013.

[PGG15a] F. Parise, B. Gentile, S. Grammatico, and J. Lygeros, *Network aggregative games: Distributed convergence to Nash equilibria*, Proceedings of the IEEE Conference on Decision and Control (2015), pp. 2295–2300.

[PGG15b] F. Parise, S. Grammatico, B. Gentile, and J. Lygeros, *Network aggregative games and distributed mean field control via consensus theory*, arXiv preprint arXiv:1506.07719 (2015).

[PGL17] F. Parise*, B. Gentile*, and J. Lygeros, *A distributed algorithm for average aggregative games with coupling constraints*, arXiv preprint arXiv:1706.04634 (2017).

[PGP16] D. Paccagnan*, B. Gentile*, F. Parise*, M. Kamgarpour, and J. Lygeros, *Distributed computation of generalized Nash equilibria in quadratic aggregative games with affine coupling constraints*, Proceedings of the IEEE Conference on Decision and Control (2016), IEEE, pp. 6123–6128.

[PKL16] D. Paccagnan, M. Kamgarpour, and J. Lygeros, *On aggregative and mean field games with applications to electricity markets*, Control Conference (ECC), 2016 European, IEEE, pp. 196–201.

[PP09] Y. Pan and L. Pavel, *Games with coupled propagated constraints in optical networks with multi-link topologies*, Automatica **45** (2009), no. 4, 871–880.

[PR03] M. Patriksson and R. T. Rockafellar, *Sensitivity analysis of aggregated variational inequality problems, with application to traffic equilibria*, Transportation Science **37** (2003), no. 1, 56–68.

[PSP10] J. Pang, G. Scutari, D. P. Palomar, and F. Facchinei, *Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach*, IEEE Transactions on Signal Processing **58** (2010), no. 6, 3251–3271.

[QM89] Y. Qiu and T. L. Magnanti, *Sensitivity analysis for variational inequalities defined on polyhedral sets*, Mathematics of Operations Research **14** (1989), no. 3, 410–432.

[Ros65] J. B. Rosen, *Existence and uniqueness of equilibrium points for concave N-person games*, Econometrica: Journal of the Econometric Society (1965), 520–534.

[RT02] T. Roughgarden and É. Tardos, *How bad is selfish routing?*, Journal of the ACM (JACM) **49** (2002), no. 2, 236–259.

[San75] A. Sandmo, *Optimal taxation in the presence of externalities*, The Swedish Journal of Economics (1975), 86–98.

[San01] W. H. Sandholm, *Potential games with continuous player sets*, Journal of Economic Theory **97** (2001), no. 1, 81–108.

[San10] ———, *Population games and evolutionary dynamics*, The MIT press, 2010.

[She85] Y. Sheffi, *Urban transportation network*, Prentice Hall (1985).

[Smi79] M. J. Smith, *The existence, uniqueness and stability of traffic equilibria*, Transportation Research Part B: Methodological **13** (1979), no. 4, 295–304.

[SPF10] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, *Convex optimization, game theory, and variational inequality theory*, IEEE Signal Processing Magazine **27** (2010), no. 3, 35–49.

[SPF12] ———, *Monotone games for cognitive radio systems*, Distributed Decision-Making and Control (2012).

[SW79] G. Salinetti and R. J.-B. Wets, *On the convergence of sequences of convex sets in finite dimensions*, Siam review **21** (1979), no. 1, 18–33.

[Tob86] R. L. Tobin, *Sensitivity analysis for variational inequalities*, Journal of Optimization Theory and Applications **48** (1986), no. 1, 191–204.

[Tob90] ———, *Sensitivity analysis for a Cournot equilibrium*, Operations research letters **9** (1990), no. 5, 345–351.

[VGA15] M. G. Vayá, S. Grammatico, G. Andersson, and J. Lygeros, *On the price of being selfish in large populations of plug-in electric vehicles*, Proceedings of the IEEE Conference on Decision and Control (2015), IEEE, pp. 6542–6547.

[Vic69] W. S. Vickrey, *Congestion theory and transport investment*, The American Economic Review **59** (1969), no. 2, 251–260.

[VNM45] J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior*, Princeton University Press Princeton, NJ, (1945).

[War52] J. G. Wardrop, *Some theoretical aspects of road traffic research.*, Proceedings of the institution of civil engineers **1** (1952), no. 3, 325–362.

[Wik17] Wikipedia, *Mean value theorem — wikipedia, the free encyclopedia*, 2017, [Online; accessed 1-December-2017].

[Yen95] N. D. Yen, *Lipschitz continuity of solutions of variational inequalities with a parametric polyhedral constraint*, Mathematics of Operations Research **20** (1995), no. 3, 695–708.

[YP17] P. Yi and L. Pavel, *A distributed primal-dual algorithm for computation of generalized Nash equilibria with shared affine coupling constraints via operator splitting methods*, arXiv preprint arXiv:1703.05388 (2017).

[YSM11] H. Yin, U. V. Shanbhag, and P. G. Mehta, *Nash equilibrium problems with scaled congestion costs and shared constraints*, IEEE Transactions on Automatic Control **56** (2011), no. 7, 1702–1708.

[ZCF17] L. Zino, G. Como, and F. Fagnani, *On imitation dynamics in potential population games*, arXiv preprint arXiv:1709.04748 (2017).

[ZF16] M. Zhua and E. Frazzoli, *Distributed robust adaptive equilibrium computation for generalized convex games*, Automatica **63** (2016), 82–91.

[ZWH15] Y. Zhou, M. Wang, H. Hao, L. Johnson, and H. Wang, *Plug-in electric vehicle market penetration and incentives: a global review*, Mitigation and Adaptation Strategies for Global Change **20** (2015), no. 5, 777–795.