CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO
DEPARTAMENTO DE MATEMÁTICAS

TEMAS EN TRANSPORTE ÓPTIMO

T E S I S

QUE PRESENTA

SERGIO JOSHAFAT GÓMEZ MÉNDEZ

PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS

EN LA ESPECIALIDAD DE
MATEMÁTICAS

DIRECTOR DE TESIS:
DR. ONÉSIMO HERNÁNDEZ LERMA

CIUDAD DE MÉXICO. NOVIEMBRE, 2019

CENTER FOR RESEARCH AND ADVANCED STUDIES OF THE NATIONAL POLYTECHNIC INSTITUTE

CAMPUS ZACATENCO
DEPARTMENT OF MATHEMATICS

TOPICS ON OPTIMAL TRANSPORT

T H E S I S

A DISSERTATION PRESENTED BY

SERGIO JOSHAFAT GÓMEZ MÉNDEZ

TO OBTAIN THE DEGREE OF
MASTER IN SCIENCE

IN THE SPECIALITY OF
MATHEMATICS

THESIS ADVISOR:
DR. ONÉSIMO HERNÁNDEZ LERMA

CIUDAD DE MÉXICO. NOVIEMBRE, 2019

To my encouraging father,
my brave mother,
my strong grandparents
and to my beloved Paulina.
For everything that has been and will be.

Agradecimientos

A mi padre por enseñarme que el esfuerzo y el cariño no conocen de límites. A mi madre por darme la fuerza para seguir adelante y enseñarme que la vida se enfrente con coraje y humildad. A mi abuelo Luis por enseñarme a ser disciplinado y constante en la vida. A mi cuquita por consentirme y amarme de forma inmensurable. A mi abuelo Jorge por sus consejos y su respeto. A mi abuela Rosalia, por su amor y sus oraciones. A mi amada Paulina por inspirarme a ser mejor persona cada día, por todo su amor y su apoyo incondicional.

A cada uno de los miembros de mi familia, en especial a Luis Omar, Claudia, Eduardo, Patricia, Andrea, Carolina, Dana, Diego, Katia, Paulina, Rocío y Michelle. A la señora Claudia, el señor Isidro, Cinthia, Katia y mi querida Pamela. A todos mis viejos y nuevos amigos, en especial a Diego, Oscar Camacho, Agustín, Oscar Méndez, Ángel, Sharon Lopez, David Hatch, Alejandra, Jorge Mancera, Nallely, Itzel, Cosijoeza, Ilse, Hugo, Isaac, Aaron, Luis Ladd, Jesica, Jose, Andrea, Christopher, Pegueros, María Lechuga, Jorge Santos, Julio, Saul, Lidia, Jennifer, Pablo, Yuriko, Wincy, Alejandra, Nelly, Daniela Varón y al mejor amigo que jamás podré terminar de agradecer; a mi padre. A todos los que no menciono pero están presentes en mí. Gracias por todo: aventuras, consejos, pláticas, conocimiento, diversión y cariño. Por todo lo que ha sido y será.

En especial quiero agradecer al Dr. Onésimo Hernández-Lerma por sus enseñanzas, su orientación y apoyo al dirigir mi tesis, su paciencia y, sobre todo, por darme la oportunidad de trabajar bajo su tutela. Jamás podré retribuir todo lo que he aprendido de él y de mis sinodales.

Finalmente, agradezco al CONACYT por el financiamiento de esta tesis.

Abstract

Optimal transport theory plays nowadays an important role in optimisation– finite dimensional or infinite dimensional–. In this thesis we study three topics related to optimal transport. To set the stage, we introduce the Monge-Kantorovich problem, and give sufficient conditions for the solvability of the problem, and several characteristics and implication associated to the solution: First, we focused in static *potential games* with infinitely-many players, combining the ideas presented in the works of Robert J. Aumann, and Morderer & Shapley. We outline the results due to Adrien Blanchet and Guillaume Carlier, associating an optimal transport plan to a potential function, where it is proved that minimizers of the potential are Cournot-Nash equilibria for such game. We provide a detailed proof that –for some sequence of classical N-Games– the limit of Nash equilibria converges to a Cournot-Nash equilibria, when the number of players tends to infinity; second, we deduce that the optimal trajectories of a set of discrete-time non-linear control systems are given by sequences of optimal transport plans associated to a non linear map; finally, we briefly sketch the links between population games and *displacement interpolation*–the time-depended version of optimal transport–, that provides a natural and intrinsic characterisation of the population dynamics associated to a gradient flow over a space of probabilities.

Resumen

En esta tesis estudiamos tres temas relacionados al transporte óptimo. Para establecer las bases de la teoría con la cual estaremos trabajando, introducimos el problema de Monge-Kantorovich y damos condiciones suficientes para la solución de este, además de diferentes caracterizaciones asociadas a dicha solución: Primero, estudiamos la relación de la teoría del transporte con juegos estáticos no cooperativos cuando el número de jugadores es infinito, esto, basándonos en las ideas presentadas en los trabajos de Robert J. Aumann y Monderer & Shapley. Detallamos los resultados de Adrien Blanchet y Guillaume Carlier acerca de cómo asociar un transporte óptimo a una función potencial, donde se prueba que los minimizadores de dicha función son equilibrios de Cournot Nash para el juego. Damos una prueba detallada acerca de cómo ciertas sucesiones de equilibrios de Nash, en el sentido clásico, convergen a equilibrios de Cournot-Nash, cuando el número de jugadores tiende a infinito; Segundo, deducimos que las trayectorias óptimas de conjuntos de sistemas de control no lineales a tiempo discreto están caracterizadas por planes de transporte óptimo asociados a un mapa no lineal; finalmente, damos un bosquejo de los vínculos entre juegos poblacionales y la versión del problema del transporte dependiente del tiempo, inter-polación desplazada, la cual provee una caracterización natural e intrínseca de cómo relacionar dinámicas poblacionales con flujos gradientes sobre espacios de probabilidad.

Contents

Notation

The following is some notation and terminology used throughout the thesis.

Id will denote the identity mapping, whatever the space. If A is a subset of a space $\mathbf{X}$, then the function I_A is the indicator function of the set A: $I_A(x) = 1$ if $x \in A$, and 0 otherwise.

The functions $\Pi_{\mathbf{X}} : \mathbf{X} \times \mathbf{Y} \to \mathbf{X}$ and $\Pi_{\mathbf{Y}} : \mathbf{X} \times \mathbf{Y} \to \mathbf{Y}$ stand for the projection functions over the spaces $\mathbf{X}$ and $\mathbf{Y}$, respectively, defined as $\Pi_{\mathbf{X}}(x, y) = x$ and $\Pi_{\mathbf{Y}}(x, y) = y$ for each $(x, y) \in \mathbf{X} \times \mathbf{Y}$.

$\mathbb{N}$ is the set of positive integers. A sequence of elements in $\mathbf{X}$ is written $\{x_n\}_{k\in\mathbb{N}}$ or $\{x_n\}_{k=1}^{\infty}$. $\mathbb{R}$ is the set of real numbers, and $\mathbb{R}_+$ denote the set of non-negative real numbers, i.e., $[0, \infty)$. $\mathbb{R}^n$ is the $n-$dimensional Euclidean space.

Let $\mathbf{X}$ be a topological space. Then $C(\mathbf{X})$ and $C_b(\mathbf{X})$ denote the set of all continuous and real valued functions, and the set of all bounded continuous real valued functions, respectively. We set $\mathcal{C}_t(\mathbf{X})$, as the set of differentiable functions indexed by t, with value in $\mathbf{X}$.

Let $(\mathbf{X}, \mathcal{F})$ be measurable space, with $\mathbf{X}$ a topological space, and $\mathcal{F}$ the corresponding $\sigma-$algebra. Then:

- the space of probability measures on $\mathbf{X}$ is denoted by $\mathscr{P}(\mathbf{X})$, the space of all finite measures is denoted by $\mathcal{M}_+(\mathbf{X})$, and the set of all signed measures is denoted by $\mathcal{M}(\mathbf{X})$;

- if μ is a measure on $\mathbf{X}$, then a subset $A \subset \mathbf{X}$ is said to be $\mu-$negligible, if there exist a set $B \in \mathcal{F}$ such that $A \subset B$ and $\mu(B) = 0$;

- a measure μ on $\mathbf{X}$ is said to be concentrated on C if $\mathbf{X} \setminus C$ is $\mu-$negligible;

- the smallest closed set where μ is concentrated, is called the support of μ and we write $\text{Spt}\,\{\mu\}$.

- if $(\mathbf{Y}, \mathcal{G})$ is a Borel measurable space,μ is a Borel measure on $\mathbf{X}$ and $T : \mathbf{X} \to \mathbf{Y}$ is a Borel map, then $T_\#\mu$ stands for the image measure of μ by T, defined by

$$(T_\#\mu)[A] := \mu[T^{-1}(A)],$$

  for every $A \in \mathcal{G}$;

- if $(\Omega, \mathcal{O}, \mathbb{P})$ is a probability space, and $X : \Omega \to \mathbb{R}$ a random variable, then the measure induce by X (also known by *law* of X) is $X_{\#}\mathbb{P}$. Thar is $(X_{\#}\mathbb{P})[A] = \mathbb{P}(X^{-1}(A))$, $A \in \mathcal{B}(\mathbb{R})$

In particular, we will assume throughout the thesis that $\mathscr{P}(\mathbf{X})$ comes from a *Polish space* $\mathbf{X}$, which is a complete, separable metric space, equipped with the Borel $\sigma-$algebra.

Introduction

Optimal transport –also known as the mass transportation problem– is nowadays one of the most prolific areas in mathematics. It was originally introduced by G. Monge [51] in 1781 and was reformulated as a linear programming problem by Kantorovich [31] in 1942. It originated in an old mathematical optimisation problems as an allocation of resources problem between mines and factories, and the problem of digging and filling pits [37]. It encompasses several theoretical areas such as real analysis [49, 27], functional analysis [47], Euclidean geometry [37, 31]. and differential topology [1, 49, 43]. Currently, it embraces a wide range of applications to, let us say, econometrics [26, 25], evolutionary dynamics [40], networks optimisation [14, 10, 24, 13, 51], risk theory [9], machine learning [48], etc.

In this work we are particularly concerned in the implementation of optimal transport in problems involving systems with an infinity of agents and dynamical systems. In specific, systems with infinitely-many agents have occupied a distinguished position in economics and game theory during the last five decades, since Robert J. Aumann proposed a model of competitive markets with a continuum of traders [2, 3]; and, Larsry and Lyons [34] introduced the notion of *mean field games* that are somewhat reminiscent of the classical mean field approaches in statistical mechanics and physics or in quantum mechanics and quantum chemistry. Several ideas and results developed in these works have established the framework for several contemporary research topics which we have taken as a starting point. For a general approach to mean field techniques in physics see [45].

In order to illustrate the relevance of the theory, let us consider the *holidays choice problem*: suppose that we wish to analyse a system with a huge amount of agents, each of them is supposed to behave rationally, in other words, each agent seeks welfare which is not supposed to be dependent only in the strategy taken by itself, but the strategies taken by the other agents. Likewise, suppose that if one agent changes his/her strategy then the outcomes of the system will not be affected at all, thus, the actions taken by a "small" set of agents are negligible to the system. For example, consider a population of agents whose location is distributed according to some Borel probability measure μ over a compact space $\mathbf{X} \subset \mathbb{R}^2$. These agents have to choose their holiday destination among all possible destinations in the set $\mathbf{Y}$, that is also a compact space. The commuting cost from x to y is given by $c(x, y)$. In addition to the commuting cost, agents incur costs resulting from interactions with other agents represented by the map $\nu \mapsto \mathcal{V}[\nu]$, where ν will represent the distribution of the population over all holiday locations, that could encode the aversion to congestion, in other words, the more crowded location results in more disutility for the agents. Some interesting questions could be: How does the population distributed among all destination places? What is the average cost for the population? How does the population behave if we take a time-dependent version of this problem? Later on, based on the results of A. Blanchet[8], we will relate the *holidays choice* problem with the optimal transport problem.

In Chapter 2, – based on the works of F. Santambrogio [47], C. Villani [49] and I. Ekeland [22] – we formally introduce the optimal transport problem and give a global characterisation of the solution.

The study of games mainly relies on finding equilibrium structures to determine optimal strategies for the players. Several structures of equilibria have been explored in games, in order to get closer to reality. Several example of different approaches in game theory could be found in literature, for instance we refer to [50, 38, 39].

"An equilibrium is not always an optimum; it might not even be good. This may be the most important discovery of game theory."

Ivar Ekeland (2016)

In Chapter 3, we look more closely to the *holidays choice* problem: we review some basic aspects of game theory; we describe the general problem of games with infinitely many players, and finally, we outline the proofs of the main results: [8, Theorem 3.3] establishes that under certain assumptions the notion of equilibrium is directly related to and characterised by the notion of optimal transport (see lemma 2.2.4), and [7, Theorem 4.1], which concludes that the notion of equilibria in games with a continuous set of players can be obtained as a limit of Nash equilibria of games with N players (see 2.1.1 for the formal definition of Nash equilibria), when N tends to infinity (see Theorems 2.3.1 and 2.3.2). These notions of limits of Nash equilibria are based on the extrapolation of mean field techniques in the seminal work of mean field games of Cardialaguet [11, Section 2], and also in the work of A. Blanchet and G. Carlier [7].

"The study of a n-person game for which the accepted ethics of fair play imply non-cooperative playing is, of course, an obvious application for game theory. The complexity of the mathematical work needed for a complete investigation increases rather rapidly, however, with increasing complexity of the game; so only be feasible using approximate computational methods"

John Nash [38, p.294]

Despite there is a fundamental existence theorem for Nash-equilibria [39], in practice, the calculation of those equilibria can be a very challenging task, even for computers. However, the study of potential games has established a framework for finding those equilibria by solving one optimisation problem associated to a potential function [36]. Thus, in Chapter 4, we study certain classes of discrete-time *optimal control* problems. The aim of the chapter is to give a first notion of dynamical systems and how optimal transport naturally arises at describing how the optimal controls must behave, and to find Nash-equilibria for a game associated to the control problem. We describe when a system can be "controlled" (see Proposition 3.1.3), and we show that an optimal control exists (see Theorem 3.2.1) and is associated to a sequence of optimal transport problems. This chapter is based on the work of K. Elamvazhuthi, P. Grover, and S. Berman [23].

In Chapter 5, we establish the relation between the theory of *population games* with a time-dependent version of optimal transport known as *displacement interpolation* [49, Chapter 7]. We study population games with a discrete strategy set in such a way that the evolutionary (mean) dynamics on the discrete set possesses the same connection as that of mean field games and displacement interpolation: we describe the state of the game with a probability measure over the set of strategies, and then, we endow the probability space with a Wasserstein metric (see Definitions 1.2.1 and 4.1.3), in such a way that the space can be seen as a Riemannian manifold [1]. It can be proved that the evolutionary dynamics can be described by a Fokker-Planck equation and also can be viewed as a gradient flow of free energies in the probability Riemannian manifold [40, 17, 16, 52]. However, this topic exceeds the scope of this thesis, therefore, for the proof of theorem 4.1.4 we refer the reader to [18].

Finally, to make this thesis as self-contained as possible, Appendix A contains a summary of some results of real analysis, functional analysis, and other areas of mathematics.

Chapter 1

Optimal Transport

The aim of this chapter is to set the stage for the rest of the thesis by formally introducing the optimal transport problem. Some main results about the qualitative picture of the solution to the mass transportation problem are accompanied by their proofs, but some other proofs exceed the scope of this thesis. Therefore, we refer C. Villani [49], F. Santambrogio [47], González-Hernández et al. [28] and Ivar Ekeland [22] for the bigger picture about optimal transport theory.

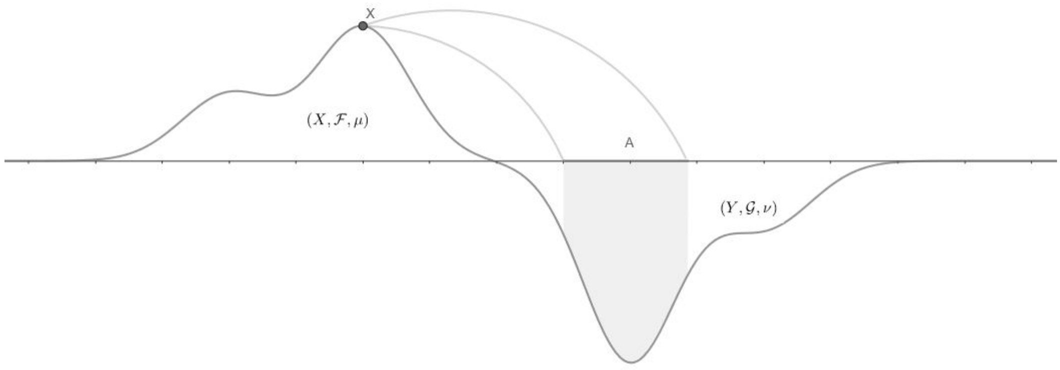
In 1781 – as we have mentioned in the introduction– Gaspard Monge introduced the problem of mass transportation; he asked himself about the optimal way to transport certain amount of a homogeneous good from N source points in the set $\mathbf{X}$, to M destination points in the set $\mathbf{Y}$, minimising the average distance traveled. In this case Monge supposed that the demand is finite and equal to the production. For example, $\mathbf{X}$ could be the set of mines near a village, and $\mathbf{Y}$ could be the set of forges in the village, then, the total production (and demand) among all mines (respectively forges) is given by a discrete probability density μ (respectively ν) with support in the set of mines (respectively forges). Here $\mu(\{j\})$ represents the proportion of the demand that is satisfied by the mine $j \in \mathbf{X}$ and $\nu(\{k\})$ the proportion of the production demanded by the forge $k \in \mathbf{Y}$, and $d_{j,k}$ represents the distance between the mine j and the forge k. We look for a matrix $\pi = [\pi_{j,k}]_{j=1,k=1}^{N,M}$, – where $\pi_{j,k}$ represents the proportion demanded by forge k, covered by mine j– in such a way that $\sum_{j\in\mathbf{X}} \pi_{j,k} = \nu(\{k\})$, for all $k \in \mathbf{Y}$, $\sum_{k\in\mathbf{Y}} \pi_{j,k} = \mu(\{j\})$, for all $j \in \mathbf{X}$, and $\sum_{j\in\mathbf{X},k\in\mathbf{Y}} \pi_{j,k} d_{j,k}$ is as small as possible. Note that this is a linear problem, and can be solved by any method of linear programming. It was firstly introduced as a linear programming problem by Kantorovich [31].

The restriction over the matrix π inspired the next definition:

Definition 1.0.1. *Let* $(\mathbf{X}, \mathcal{F}, \mu)$ *and* $(\mathbf{Y}, \mathcal{G}, \nu)$ *be two probability spaces. We say that two random variables* X *and* Y*, on some probability space* $(\mathbf{Z}, \mathcal{H}, \pi)$*, is a coupling if law*$(X) = \mu$ *and law*$(Y) = \nu$*. We also say that the pair* (X, Y) *is a coupling of* (μ, ν)*. Additionally, we define* $\Pi(\mu, \nu)$ *as the set of* $\pi \in \mathscr{P}(\mathbf{X} \times \mathbf{Y})$ *whose marginals are* μ *and* ν*, respectively.*

Remark. *If* μ *and* ν *are the only measures involved in the problem, we can consider* $\mathbf{Z} = \mathbf{X} \times \mathbf{Y}$ *and* $\mathcal{H} = \mathcal{F} \otimes \mathcal{G}$*, then the set of couplings is equivalent to* $\Pi(\mu, \nu)$*.*

Figure 1.1: The conditioned measure by the point $x \in \mathbf{X}$, given a coupling (X, Y) for (μ, ν)



We could interpret this as follows: if $A \in \mathcal{G}$ and $x \in \mathbf{X}$, then $\pi(Y \in A|X = x)$ represents the probability of transporting a homogeneous unit of mass from one source point x to a measurable subset A.

Based on these definitions, we can observe that the existence, or the absence, of the coupling for any pair of probability spaces must depend on several characteristics of such spaces. So, firstly, let us suppose that all the information about the value of the random variable Y is given by the value of the random variable X, i.e., Y is a function of X. This motivates the following definition:

Definition 1.0.2. *A coupling (X, Y) for (μ, ν) is said to be deterministic if there exist a measurable function $T : \mathbf{X} \to \mathbf{Y}$ such that $Y = T(X)$.*

Informally, one can say that T transports the mass given by the measure μ, to the mass given by the measure ν, so T is commonly called a *transport map.*

Let us suppose that there exists a *cost* function $c : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}$ that can be interpreted as the work needed to move one unit of mass from location x to location y – just as in the example of mines and forges –. Let us also suppose that there exists at least one coupling $\pi \in \Pi(\mu, \nu)$, then the cost associated to that coupling is given by

$$\int_{\mathbf{X}\times\mathbf{Y}} c(x, y) d\pi(x, y).$$

Naturally, we aim to find a coupling for (μ, ν) that minimises the total transportation cost given by c.

Definition 1.0.3. *Let $(\mathbf{X}, \mathcal{F}, \mu)$, $(\mathbf{Y}, \mathcal{G}, \nu)$ be two probability spaces, and $c : \mathbf{X}\times\mathbf{Y} \to$*

$\mathbb{R}$ *the cost function. The Monge-Kantorovich minimisation problem is given by*

$$\begin{array}{ll} \textit{Minimise} & \displaystyle\int_{\mathbf{X}\times\mathbf{Y}} c(x,y)d\pi(x,y) \\ \textit{subject to} & \pi \in \Pi(\mu,\nu) \end{array}$$

All couplings of (μ, ν) *are called transference plans, and those achieving the infimum are called optimal transference plans.*

Consider the map $\pi \mapsto C_c(\pi) := \int_{\mathbf{X}\times\mathbf{Y}} cd\pi$, where $c : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}$ is a fixed measurable function. In particular if c is continuous, then, by definition of weak convergence, C_c – also written as a bi-linear form $\langle c, \pi\rangle$ – is continuous over the space $\mathscr{P}(\mathbf{X} \times \mathbf{Y})$. We are specially interested in the properties of C_c when c is as below:

Definition 1.0.4 (Semi-continuity)**.** *If* $\mathbf{Z}$ *is metric space. A function* $c : \mathbf{Z} \to \mathbb{R} \cup \{\infty\}$ *is said to be lower semi-continuous if*

$$\liminf_{z\to z_0} c(z) \geq c(z_0)$$

for each $z_0 \in \mathbf{Z}$*, and upper semi-continuous if*

$$\limsup_{z\to z_0} c(z) \leq c(z_0)$$

for each $z_0 \in \mathbf{Z}$*.*

Remark. *Is well-known that* c *is lower semi-continuous if and only if the set* $\{z \in \mathbf{Z} : c(z) \leq \alpha\}$ *is closed for each* $\alpha \in \mathbb{R}$*.*

Lemma 1.0.5. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two Polish spaces, and* $c : \mathbf{X} \times \mathbf{Y} \to \mathbb{R} \cup \{\infty\}$ *be a nonnegative lower semi-continuous cost function. Then the mapping* $\pi \mapsto C_c(\pi)$ *is lower semi-continuous with respect to the weak topology.*

Proof. By lemma A.0.4 in Appendix A, c can be written as the point-wise limit of a non-decreasing sequence $\{c_n\}_n^\infty$ of continuous bounded functions. By monotone convergence,

$$\int_{\mathbf{X}\times\mathbf{Y}} cd\pi = \lim_{n\to\infty} \int_{\mathbf{X}\times\mathbf{Y}} c_n d\pi.$$

As for each $n \in \mathbb{N}$, $c_n \leq c$ and c_n is bounded and continuous, then, by definition of weak convergence

$$\int_{\mathbf{X}\times\mathbf{Y}} cd\pi \leq \lim_{k\to\infty} \left(\lim_{n\to\infty} \int_{\mathbf{X}\times\mathbf{Y}} c_n d\pi_k \right) = \liminf_{k\to\infty} \int_{\mathbf{X}\times\mathbf{Y}} cd\pi_k.$$

Hence, the map $\pi \mapsto C_c(\pi)$ is lower semi-continuous. □

Remark. *Note that if* $c : \mathbf{X}\times\mathbf{Y} \to \mathbb{R}$ *is lower semi-continuous, and is bounded below by an upper semi-continuous function* $\hat{c}$*. Then the function* $c - \hat{c}$ *is a nonnegative lower semi-continuous function, then Lemma 1.0.5 applies.*

It just remains to introduce the concept of tightness in order to solve the Monge-Kantorovich problem.

Definition 1.0.6. *A subset $Q \subset \mathscr{P}(\mathbf{X})$ is said to be tight if, for every $\epsilon > 0$, there exists a compact subset $K \subset \mathbf{X}$ such that $\mu(\mathbf{X} \setminus K) < \epsilon$ for every $\mu \in Q$.*

Lemma 1.0.7 (Tightness of the transference plans)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be two Polish spaces. Let $P \subset \mathscr{P}(\mathbf{X})$ and $Q \subset \mathscr{P}(\mathbf{Y})$ be tight, as in Definition 1.0.6. Then the set $\Pi(P, Q) := \{\pi \in \mathscr{P}(\mathbf{X} \times \mathbf{Y}) : \Pi_{\mathbf{X}}\#\pi \in P, \Pi_{\mathbf{Y}}\#\pi \in Q\}$ is tight in $\mathscr{P}(\mathbf{X} \times \mathbf{Y})$.*

Proof. As P and Q are tight, for each $\epsilon > 0$, there exist compact sets K_ϵ and L_ϵ in $\mathbf{X}$ and $\mathbf{Y}$, respectively, such that

$$(\forall \mu \in P) \qquad \mu(\mathbf{X} \setminus K_\epsilon) < \frac{\epsilon}{2} \qquad \textit{and} \qquad (\forall \nu \in Q) \qquad \nu(\mathbf{Y} \setminus L_\epsilon) < \frac{\epsilon}{2}.$$

Then

$$(\forall \pi \in \Pi(P, Q)) \qquad \pi(\mathbf{X} \times \mathbf{Y} \setminus K_\epsilon \times L_\epsilon) \leq \mu(\mathbf{X} \setminus K_\epsilon) + \nu(\mathbf{Y} \setminus L_\epsilon) < \epsilon.$$

This complete the proof. □

"The first good thing about optimal couplings, is that they exist."

C. Villani [49, Chapter 4]

Theorem 1.0.8. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Polish space, $\mu \in \mathscr{P}(\mathbf{X})$, $\nu \in \mathscr{P}(\mathbf{Y})$ and $c : \mathbf{X} \times \mathbf{Y} \to [0, \infty)$ lower semi-continuous. Then the Monge-Kantorovich minimisation problem admits a solution.*

Proof. Since $\mathbf{X}$ and $\mathbf{Y}$ are Polish spaces, μ and ν are regular measures, then tight. By Lemma 1.0.7 $\Pi(\mu, \nu)$ is tight, and by Prohorov's theorem [6] is also a pre-compact set. Now, let us take a sequence $\{\pi_n\}_{n=1}^{\infty} \subset \Pi(\mu, \nu)$, weakly converging to $\pi \in \mathscr{P}(\mathbf{X}, \mathbf{Y})$, and take an arbitrary $\phi \in C_b(\mathbf{X})$. Then we have for each $k \in \mathbb{N}$ that

$$\int_{\mathbf{X} \times \mathbf{Y}} \phi(x) d\pi_{n_k}(x, y) = \int_{\mathbf{X}} \phi d\mu$$

and passing to the limit when $k \to \infty$ we have

$$\int_{\mathbf{X} \times \mathbf{Y}} \phi(x) d\pi(x, y) = \int_{\mathbf{X}} \phi d\mu.$$

Since $\phi \in C_b(\mathbf{X})$ was arbitrary, it follows that $\Pi_{\mathbf{X}}\#\pi = \mu$. Analogously $\Pi_{\mathbf{Y}}\#\pi = \nu$. Therefore $\Pi(\mu, \nu)$ in fact is a compact set. Finally, by Lemma 1.0.5, we have a lower semi-continuous function $\pi \mapsto C_c(\pi)$ on a compact set, the desired conclusion follows from Theorem A.0.5

□

1.1 Duality

Kantorovich, in his seminal work "*On the Translocation of Masses*" [31], expressed the problem of optimal transport as an infinite dimensional linear programming problem. Moreover, he applies the concept of duality to the problem and defines the potentials –so-called Kantorovich potentials–for the Monge-Kantorovich problem.

In this section we introduce the Dual Monge-Kantorovich problem, we give sufficient conditions for the solvability of the problem and we discuss some geometrical properties of the problem.

Definition 1.1.1. *Define $\mathbb{B}(\mathbf{X}, \mathbf{Y})$ as the set of all bijection between $\mathbf{X}$ and $\mathbf{Y}$. We say that $\xi \in \mathbb{B}(\mathbf{X}, \mathbf{Y})$ is optimal for c if*

$$(\forall \xi' \in \mathbb{B}(\mathbf{X}, \mathbf{Y})) \qquad \sum_{x\in\mathbf{X}} c(x, \xi(x)) \le \sum_{x\in\mathbf{X}} c(x, \xi'(x)).$$

Proposition 1.1.2. *If $\mathbb{B}(\mathbf{X}, \mathbf{Y})$ is nonempty and finite, then $\xi \in \mathbb{B}(\mathbf{X}, \mathbf{Y})$ is optimal for c if and only if, for every set, $\{x_0, \dots, x_N = x_0\} \subset \mathbf{X}$, $n \in \mathbb{N}$, we have:*

$$\sum_{n=0}^{N-1} [c(x_n, \xi(x_n)) - c(x_n, \xi(x_{n+1}))] \ge 0. \tag{1.1.1}$$

Proof. (Sufficiency)
Let us take $\xi \in \mathbb{B}(\mathbf{X}, \mathbf{Y})$ satisfying (1.1.1), and $\zeta \in \mathbb{B}(\mathbf{X}, \mathbf{Y})$ arbitrary. Choosing an arbitrary $x_0 \in \mathbf{X}$ and denoting $y_1 := \zeta(x_0)$. There is a unique $x_1 \in \mathbf{X}$ such that $\xi(x_1) = y_1$. We denote $y_2 := \zeta(x_1)$. There is a unique $x_2 \in \mathbf{X}$ such that $\xi(x_2) = y_2 = \zeta(x_1)$. Inductively we denote $y_{n+1} := \zeta(x_n)$, and $x_n \in \mathbf{X}$, such that $\xi(x_{n+1}) = \zeta(x_n)$. As there are a finite number of elements, we can take $N := \inf\{n \in \mathbb{N} : x_0 = x_N\}$. Note that any other pair of elements in $S_1 := \{x_1, \dots, x_N\}$ is equal but x_0 and x_N. Then, by hypothesis, we have

$$\sum_{n=0}^{N-1} [c(x_n, \xi(x_n)) - c(x_n, \xi(x_{n+1}))] = \sum_{x\in S_1} [c(x, \xi(x)) - c(x, \zeta(x))] \ge 0.$$

If $S_1 = \mathbf{X}$, then we have finished. If not, let us take $x_0' \in \mathbf{X} \setminus S_1$, and repeat the procedure to obtain S_2. Let us take an arbitrary $x \in S_1$, then there is an integer m such that $x_m = x$, using the above procedure and his uniqueness $N - m$ times, we got the same last $N - m$ element in S_1, then as $x_N = x_0$ we continue the procedure. If we apply the same procedure for any $x \in S_1$, then we will get the same set S_1. Thus, S_1 and S_2 are disjoint, if $S_1 \cup S_2 \ne \mathbf{X}$, then we continue. As there are as most disjoint sets $S_1, S_2, \dots,$ as elements in $\mathbf{X}$, $\mathbf{X} = \cup_{k=1}^{K} S_k$, for some $K \in \mathbb{N}$. Therefore

$$\sum_{k=1}^{K} \sum_{x\in S_k} [c(x, \xi(x)) - c(x, \zeta(x))] = \sum_{x\in\mathbf{X}} [c(x, \xi(x)) - c(x, \zeta(x))] \ge 0.$$

(Necessity)
Let $\xi \in \mathbb{B}(\mathbf{X}, \mathbf{Y})$ be optimal, and let $S = \{x_0, \dots x_N = x_0\} \subset \mathbf{X}$. Define the map

$\zeta : \mathbf{X} \to \mathbf{Y}$ such that for $0 \leq n \leq N-1$, $\zeta(x_n) := \xi(x_{n+1})$, and for $x \notin S$, $\zeta(x) := \xi(x)$. Then $\zeta \in \mathbb{B}(\mathbf{X}, \mathbf{Y})$, and using that ξ is optimal we have that

$$\sum_{n=0}^{N} c(x_n, \xi(x_n)) = \sum_{x\in S} c(x, \xi(x)) \geq \sum_{x \in S} c(x, \zeta(x)) = \sum_{n=0}^{N-1} c(x_n, \xi(x_{n+1})).$$

□

Based on this property we give the next definition, that will be of relevance in the concept of duality and will describe several characteristics for the optimal transport under suitable assumptions.

Definition 1.1.3 (Cyclical monotonicity)**.** *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be arbitrary sets, and* $c : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}$ *a given function. A subset* $\Gamma \subset \mathbf{X} \times \mathbf{Y}$ *is said to be c-cyclical monotone if, for any* $N \in \mathbb{N}$, *and any family* $(x_0, y_0) \dots (x_N, y_N)$ *of points in* Γ, *it holds that*

$$\sum_{j=0}^{N} c(x_j, y_j) \leq \sum_{j=0}^{N} c(x_j, y_{j+1}), \tag{1.1.2}$$

where $y_{N+1} := y_0$. *Also we say that a transference plan is c-cyclical monotone if it is concentrated on a c-cyclical monotone set.*

On the one hand, let us suppose that we aim to optimally transport a homogeneous good between two different sets of location, $\mathbf{X}$ and $\mathbf{Y}$. On the other hand, suppose that there is a third person who is expert in the area of transportation, and he offers to do the transportation instead of us. The expert could find cheaper ways of doing the transportation and in this case, we only are going to interact with this third person and not with the owner of the other set of source places. Let us suppose that he cost of one unit of the good in location $x \in \mathbf{X}$, is given by $\psi(x)$, after the transportation the good is sold at some location $y \in \mathbf{Y}$ at a price $\phi(y)$, so the actual cost of transportation between the locations x and y is given by $\phi(y) - \psi(x)$, instead of the original cost $c(x, y)$. As the expert needs to competitively set up prices, the next inequality holds:

$$(\forall (x, y) \in \mathbf{X} \times \mathbf{Y}) \qquad \phi(y) - \psi(x) \leq c(x, y). \tag{1.1.3}$$

As in the Monge-Kantorovich problem we aim to minimise our cost, whereas in the dual problem the expert aims to maximise his profits.

Definition 1.1.4 (Dual Monge-Kantorovich Problem)**.** *Let* $(\mathbf{X}, \mu)$, $(\mathbf{Y}, \nu)$ *be two probability spaces, and* $c : \mathbf{X}\times\mathbf{Y} \to \mathbb{R}_+$ *a cost function. The dual Monge-Kantorovich problem is*

$$\sup \left\{ \int_{\mathbf{Y}} \phi(y) d\nu(y) - \int_{\mathbf{X}} \psi(x) d\mu(x) : \;\; \psi \in \mathscr{M}(\mathbf{X}), \;\; \phi \in \mathscr{M}(\mathbf{Y}), \;\; \textit{and} \;\; \phi - \psi \leq c \right\}. \tag{1.1.4}$$

Later on, it will be imposed that $\psi \in L^1(\mu)$ and $\phi \in L^1(\nu)$.

Remark. *By monotonicity of the integral and taking $\pi \in \Pi(\mu,\nu)$, we get from* (1.1.3) *that*

$$\int_{\mathbf{X}\times\mathbf{Y}} (\phi(y)-\psi(x))d\pi(x,y) = \int_{\mathbf{Y}} \phi(y)d\nu(y) - \int_{\mathbf{X}} \psi(x)d\mu(x) \le \int_{\mathbf{X}\times\mathbf{Y}} c(x,y)d\pi(x,y).$$

Since ϕ, ψ and π are arbitrary, we have

$$\sup_{\phi-\psi\le c}\left\{\int_{\mathbf{Y}} \phi(y)d\nu(y) - \int_{\mathbf{X}} \psi(x)d\mu(x)\right\} \le \inf_{\pi\in\Pi(\mu,\nu)}\left\{\int_{\mathbf{X}\times\mathbf{Y}} c(x,y)d\pi(x,y)\right\}. \quad (1.1.5)$$

Definition 1.1.5. *For any two functions $\psi : \mathbf{X} \to \mathbb{R}$ and $\phi : \mathbf{Y} \to \mathbb{R}$ we define the $c-$transform $\psi^c : \mathbf{Y} \to \mathbb{R}$ of ψ, and the $\bar{c}-$transform $\phi^{\bar{c}} : \mathbf{X} \to \mathbb{R}$ of ϕ as*

$$(\forall y \in \mathbf{Y}) \qquad \psi^c(y) := \inf_{x\in\mathbf{X}}\{\psi(x) + c(x,y)\}, \quad (1.1.6)$$

$$(\forall x \in \mathbf{X}) \qquad \phi^{\bar{c}}(x) := \sup_{y\in\mathbf{Y}}\{\phi(y) - c(x,y)\}, \quad (1.1.7)$$

The c-sub-differential set of ψ is defined by

$$\partial_c\psi := \{(x,y) \in \mathbf{X}\times\mathbf{Y} : \quad \psi^c(y) - \psi(x) = c(x,y)\}, \quad (1.1.8)$$

and at the point $x \in \mathbf{X}$,

$$\partial_c\psi(x) := \{y \in \mathbf{Y} : \quad (x,y) \in \partial_c\psi\}. \quad (1.1.9)$$

We say that the functions ψ and ϕ are tight with respect to c if $\psi = \phi^{\bar{c}}$ and $\phi = \psi^c$.

Definition 1.1.6 ($c-$convexity)**.** *Let $(\mathbf{X},\mathcal{F},\mu)$ and $(\mathbf{Y},\mathcal{G},\nu)$ be two probability spaces, and $c : \mathbf{X}\times\mathbf{Y} \to \mathbb{R}_+$. A function $\psi : \mathbf{X} \to \mathbb{R}\cup\{\infty\}$ is said to be $c-$convex if it is not identically ∞ and there exists $\zeta : \mathbf{Y} \to \mathbb{R}\cup\{\pm\infty\}$ such that $\psi = \zeta^{\bar{c}}$. Analogously $\phi : \mathbf{Y} \to \mathbb{R}\cup\{\infty\}$ is said to be $c-$concave if its not identically $-\infty$ and there exists $\gamma : \mathbf{X} \to \mathbb{R}\cup\{\pm\infty\}$ such that $\phi = \gamma^c$.*

Proposition 1.1.7. *Let $\psi : \mathbf{X} \to \mathbb{R}\cup\{\infty\}$, then ψ is $c-$convex if and only if $\psi = (\psi^c)^{\bar{c}}$.*

Proof. (Necessity)
If ψ is $c-$convex, then there exist $\phi : \mathbf{Y} \to \mathbb{R}\cup\{\infty\}$ such that $\psi = \phi^{\bar{c}}$. Let us now define

$$(\forall x,\hat{x} \in \mathbf{X})(\forall y,\hat{y} \in \mathbf{Y}) \qquad \hat{\phi}(\hat{x},\hat{y},x,y) := \phi(\hat{x}) - c(\hat{x},\hat{y}) + c(\hat{x},y) - c(x,y).$$

It is clear that $(\forall x \in \mathbf{X})$ $((\phi^{\bar{c}})^c)^{\bar{c}}(x) = \sup_{y\in\mathbf{Y}} \inf_{\hat{x}\in\mathbf{X}} \sup_{\hat{y}\in\mathbf{Y}}\{\hat{\phi}(\hat{x},\hat{y},x,y)\}$. Then if we set $\hat{y} = y$ and take the supremum, we have that $\phi^{\bar{c}} \le ((\phi^{\bar{c}})^c)^{\bar{c}}$. Then if we set $\hat{x} = x$ and take the infimum we have that $\phi^{\bar{c}} \ge ((\phi^{\bar{c}})^c)^{\bar{c}}$. Hence $\psi = (\psi^c)^{\bar{c}}$. Moreover. for any $\phi : \mathbf{Y} \to \mathbb{R}_+$, $\phi^{\bar{c}} = ((\phi^{\bar{c}})^c)^{\bar{c}}$.
(Sufficiency)
If $\psi = (\psi^c)^{\bar{c}}$, then by taking $\phi = \psi^c$, we have by definition that ψ is $c-$convex. □

Theorem 1.1.8 (Monge-Kantorovich Duality in C_b)**.** *Let* $(\mathbf{X}, \mu)$ *and* $(\mathbf{Y}, \nu)$ *be two Polish probability spaces, and let* $c : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}_+$ *be a bounded and continuous cost function. Then there is duality*

$$\begin{aligned}\min_{\pi\in\Pi(\mu,\nu)} &\int_{\mathbf{X}\times\mathbf{Y}} c(x,y)d\pi(x,y) \\ &= \sup_{(\psi,\phi)\in C_b(\mathbf{X})\times C_b(\mathbf{Y});\phi-\psi\leq c} \left(\int_{\mathbf{Y}} \phi(y)d\nu(y) - \int_{\mathbf{X}} \psi(x)d\mu(x)\right).\end{aligned}$$

Proof. First, let us take the particular case when μ and ν are measures with finite support –both of them with the same numbers of elements–, i.e., for some $n \in \mathbb{N}$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}, \qquad \nu = \frac{1}{n}\sum_{j=1}^{n} \delta_{y_j}.$$

Where $\{x_j\}_{j=1}^n \subset \mathbf{X}$ and $\{y_j\}_{j=1}^n \subset \mathbf{Y}$. In this particular case the Monge-Kantorovich problem becomes

$$\begin{array}{lll} \text{Minimise} & \sum a_{ij}c(x_i,y_j) & \\ s.t. & a_{ij} \geq 0 & (\forall i,j \in \{1,\dots,n\}), \\ & \sum_{i=1}^{n} a_{ij} = 1 & (\forall j \in \{1,\dots,n\}), \\ & \sum_{j=1}^{n} a_{ij} = 1 & (\forall i \in \{1,\dots,n\}). \end{array}$$

By Theorem 1.0.8, there exists at least one optimal transport plan

$$\pi = \frac{1}{n}\sum_{i,j=1}^{n} a_{ij}\delta_{(x_i,y_j)},$$

with support $S \subset \mathbf{X} \times \mathbf{Y}$ where $(x_i, y_j) \in S$, if and only if, $a_{i,j} > 0$. Suppose that S is not c-cyclically monotone, then there exist $N \in \mathbb{N}$ and a sequence $\{(x_{i_1}, y_{j_1}), \dots, (x_{i_N}, y_{j_N})\} \subset S$ such that

$$c(x_{i_1}, y_{j_2}) + c(x_{i_2}, y_{j_3}) + \cdots + c(x_{i_N}, y_{j_1})) < c(x_{i_1}, y_{j_1}) + \cdots + c(x_{i_N}, y_{j_N}). \quad (1.1.10)$$

Let us define $a := \min(a_{i_1,j_1}, \dots, a_{i_N,j_N})$. Define a new transference plan $\hat{\pi}$ by the formula

$$\hat{\pi} := \frac{1}{n}\sum_{i,j=1}^{n} a_{ij}\delta_{(x_i,y_j)} + \frac{a}{n}\sum_{l=1}^{N}(\delta_{(x_{i_l},y_{j_l})} - \delta_{(x_{i_l},y_{j_{l+1}})}).$$

Intuitively, the second term of the above sum redistributes a proportion a of the mass $\frac{1}{n}$ from the set $\{(x_{i_1}, y_{j_1}), \dots, (x_{i_N}, y_{j_N})\}$ to the set $\{(x_{i_1}, y_{j_2}), \dots, (x_{i_N}, y_{j_1})\} \subset \mathbf{X} \times \mathbf{Y}$. Thus, by (1.1.10), the cost associated with $\hat{\pi}$ is strictly less that the cost associated with π. This is a contradiction for the optimal transport plan π. Therefore S is c-cyclically monotone.

Now, let us suppose that μ and ν are any Borel probability measures on $(\mathbf{X}, \mathcal{F})$, respectively, $(\mathbf{Y}, \mathcal{G})$. Then there exist probability spaces $(\mathbf{\Omega_X}, \mathbb{P}_\mathbf{X})$ and $(\mathbf{\Omega_Y}, \mathbb{P}_\mathbf{Y})$, on which there are independent random variables $\{X_j\}_{j=1}^\infty$, respectively $\{Y_j\}_{j=1}^\infty$, with values in $\mathbf{X}$, respectively $\mathbf{Y}$, and law $(X_j) = \mu$, respectively law $(Y_j) = \nu$, for every $j \in \mathbb{N}$. By Theorem A.0.12, we have

$$\mu_n := \frac{1}{n}\sum_{j=1}^n \delta_{X_j(\omega_\mathbf{X})} \rightharpoonup \mu, \qquad \nu_n := \frac{1}{n}\sum_{j=1}^n \delta_{Y_j(\omega_\mathbf{Y})} \rightharpoonup \nu,$$

for $\mathbb{P}_\mathbf{X}-$almost every $\omega_\mathbf{X}$, respectively $\mathbb{P}_\mathbf{Y}-$almost every $\omega_\mathbf{Y}$, as $n \to \infty$. Note that, by Theorem A.0.9, $\{\mu_j\}_{j=1}^\infty$ and $\{\nu_j\}_{j=1}^\infty$ are tight.

For each $n \in \mathbb{N}$, let π_n be a $c-$cyclically monotone transference plan between μ_n and ν_n. By Lemma 1.0.7, $\{\pi_k\}_{k=1}^\infty$ is tight, and by Theorem A.0.9, there is a subsequence, still denoted by $\{\pi_k\}_{k=1}^\infty$, which converges weakly to some probability measure $\pi \in \Pi(\mu, \nu)$.
Let us take an arbitrary $N \in \mathbb{N}$. For each $n \in \mathbb{N}$, the product measure $\pi_n^{\otimes N} := \bigotimes_{k=1}^N \pi_n$ is concentrated on the set $\mathcal{C}(N)$ of all $((x_1, y_1), \ldots, (x_N, y_N)) \in (\mathbf{X} \times \mathbf{Y})^N$ satisfying (1.1.1). Since c is continuous, $\mathcal{C}(N)$ is a closed set, by Portmanteau Theorem [6, Section 2, Theorem 2.1], $\pi^{\otimes N}$ is also concentrated on $\mathcal{C}(N)$. Let $\Gamma = \operatorname{Spt}\{\pi\}$. Then $\Gamma^N = \operatorname{Spt}\{\pi\}^N = \operatorname{Spt}\left\{\pi^{\otimes N}\right\} \subset \mathcal{C}(N)$. Since this holds for each $N \in \mathbb{N}$, Γ is $c-$cyclically monotone.

Let us redefine $\Gamma := \operatorname{Spt}\{\pi\}$.Let us take any $(x_0, y_0) \in \Gamma$ and define

$$\begin{aligned}\psi(x) := \sup_{m\in\mathbb{N}} \sup \{&[c(x_0, y_0) - c(x_1, y_0)] + [c(x_1, y_1) - c(x_2, y_1)] \\ &+ \cdots + [c(x_m, y_m) - c(x, y_m)] : \{(x_j, y_j)\}_{j=1}^m \subset \Gamma\},\end{aligned}$$

for each $x \in \mathbf{X}$. Since ψ is the supremum of ψ_m lower semi-continuous functions, ψ is lower semi-continuous. In particular ψ is measurable. On the one hand $\psi(X_0) \geq [c(x_0, y_0) - c(x_0, y_0)] = 0$. On the other hand, by c-cyclically monotonicity of the transference plan π, $\psi(x_0)$ is the supremum over nonpositive quantities $[c(x_0, y_0) - c(x_1, y_0)] + [c(x_1, y_1) - c(x_2, y_1)] + \cdots + [c(x_m, y_m) - c(x_0, y_m)]$. Hence, $\psi(x_0) = 0$.

For each $y \in \Pi_\mathbf{Y}(\Gamma)$, denote

$$\begin{aligned}\zeta(y) := \sup \{&[c(x_0, y_0) - c(x_1, y_0)] + [c(x_1, y_1) - c(x_2, y_1)] \\ &+ \cdots + c(x_m, y) : m \in \mathbb{N}, \{(x_1, y_1), \ldots, (x_m, y)\} \subset \Gamma\},\end{aligned}$$

with $\zeta(y) = -\infty$ when $y \notin \Pi_\mathbf{Y}(\Gamma)$. Then

$$\psi(x) \sup_{y\in\mathbf{Y}}\{\zeta(y) - (x, y)\}.$$

Thus ψ is a $c-$convex function. Let $(\hat{x}, \hat{y}) \in \Gamma$. Then

$$\psi(x) \geq \sup_{m \in \mathbb{N}} \sup \{[c(x_0, y_0) - c(x_1, y_0)] + [c(x_1, y_1) - c(x_2, y_1)] \\ + \cdots + [c(\hat{x}, \hat{y}) - c(x, \hat{y})] : \{(x_j, y_j)\}_{j=1}^{m-1} \subset \Gamma\big\} = \psi(\hat{x}) + c(\hat{x}, \hat{y}) - c(x, \hat{y}).$$

In particular, $\psi(x) + (x, \hat{y}) \geq \psi(\hat{x}) + (\hat{x}, \hat{y})$. Taking the infimum over $x \in \mathbf{X}$ in the left-hand side, it follows that

$$\psi^c(\hat{y}) \geq \psi(\hat{x}) = c(\hat{x}, \hat{y}) \geq \psi^c(\hat{y}).$$

Hence, $\Gamma \subset \partial_c \psi$. Set $C := \sup_{(x,y) \in \mathbf{X} \times \mathbf{Y}} c(x, y)$. We define $\phi := \psi^c$. Since ψ is measurable, ϕ is measurable. Let $(x_0, y_0) \in \partial_c \psi$ be such that $\psi(x_0) < \infty-$ then $\phi(y_0) > -\infty$. Then, for any $c \in \mathbf{X}$, by reapplying the inequalities,

$$\psi(x) = \sup_{y \in \mathbf{Y}} \{\phi(y) - c(x, y)\} \geq \phi(y_0) - c(x, y_0) \geq \phi(y_0) - C,$$
$$\phi(y) = \inf_{x \in \mathbf{X}} \{\psi(x) + c(x, y)\} \leq \psi(x_0) + c(x_0, y) \leq \psi(x_0) + C.$$

Moreover,

$$\psi(x) = \sup_{y \in \mathbf{Y}} \{\phi(y) - c(x, y)\} \leq \sup_{y \in \mathbf{Y}} \{\psi(x_0) + C - c(x, y)\} \leq \psi(x_0) + C,$$
$$\phi(y) = \inf_{x \in \mathbf{X}} \{\psi(x) + c(x, y)\} \geq \inf_{x \in \mathbf{X}} \{\phi(y_0) - C + c(x, y)\} \geq \phi(y_0) - C.$$

So both ψ and ϕ are bounded and measurable, and therefore

$$\int_{\mathbf{X} \times \mathbf{Y}} [\phi(y) - \psi(x)] d\pi(x, y) = \int_{\mathbf{Y}} \phi(y) d\nu(y) - \int_{\mathbf{X}} \psi(x) d\mu(x)$$

is well defined. Since $\phi(y) - \psi(x) = c(x, y)$ on the support of π, it follows

$$\int_{\mathbf{Y}} \phi(y) d\nu(y) - \int_{\mathbf{X}} \psi(x) d\mu(x) = C_c(\pi).$$

Hence, there is duality. □

Remark. *Indeed, in Theorem 1.1.8, we have shown that there is strong duality, i.e., there exist* $(\psi, \phi) \in C_b(\mathbf{X}) \times C_b(\mathbf{Y})$ *such that the supremum in* (1.1.4) *is reached.*

Theorem 1.1.9 (Monge-Kantorovich Duality)**.** *Let* $(\mathbf{X}, \mu)$ *and* $(\mathbf{Y}, \nu)$ *be two Polish probability spaces, and let* $c : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}_+$ *be a lower semi-continuous cost function. Then there is duality:*

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathbf{X} \times \mathbf{Y}} c(x, y) d\pi(x, y) \\ = \sup_{(\psi, \phi) \in C_b(\mathbf{X}) \times C_b(\mathbf{Y}); \phi - \psi \leq c} \left(\int_{\mathbf{Y}} \phi(y) d\nu(y) - \int_{\mathbf{X}} \psi(x) d\mu(x) \right).$$

Proof. Since c is lower semi-continuous, by Theorem A.0.4, there exists an increasing subsequence of bounded and continuous functions $\{c_k\}_{k=0}^{\infty}$, such that $\lim_{n\to\infty} c_n(x,y) = c(x,y)$ for every $(x,y) \in \mathbf{X}\times\mathbf{Y}$. By Theorem 1.1.8, for each $k\in\mathbb{N}$, we can find $\psi_k \in C_b(\mathbf{X})$, $\phi_k \in C_b(\mathbf{Y})$ and $\pi_k \in \Pi(\mu,\nu)$, such that ψ_k is a c_k-convex function, $\phi_k = (\psi_k)^{c_k}$, and

$$\int_{\mathbf{X}\times\mathbf{Y}} c_k(x,y)d\pi_k(x,y) = \int_{\mathbf{Y}} \phi_k(y)d\nu(y) - \int_{\mathbf{X}} \psi_k(x)\mu(x).$$

Since, for each $k\in\mathbb{N}$, $\phi_k - \psi_k \le c_k \le c$, (ψ_k,ϕ_k) are admissible plans for the dual problem with cost c.

By Lemma 1.0.7, $\Pi(\mu,\nu)$ is weakly sequentially compact. Thus, up to extractions of a sequence, we can assume that $\{\pi_k\}_{k=1}^{\infty}$ converges to some $\pi\in\Pi(\mu,\nu)$. On the one hand, we have

$$\begin{aligned}\int_{\mathbf{X}\times\mathbf{Y}} c_l d\pi &= \lim_{k\to\infty}\int_{\mathbf{X}\times\mathbf{Y}} c_l d\pi_k \\ &\le \limsup_{k\to\infty}\int_{\mathbf{X}\times\mathbf{Y}} c_k d\pi_k \\ &= \limsup_{k\to\infty}\left(\int_{\mathbf{Y}}\phi_k d\nu - \int_{\mathbf{X}}\psi_k d\mu\right).\end{aligned}$$

On the other hand, by monotone convergence,

$$\int_{\mathbf{X}\times\mathbf{Y}} c d\pi = \lim_{l\to\infty}\int_{\mathbf{X}\times\mathbf{Y}} c_l d\pi.$$

Then

$$\inf_{\hat\pi\in\Pi(\mu,\nu)}\int_{\mathbf{X}\times\mathbf{Y}} c d\hat\pi \le \int_{\mathbf{X}\times\mathbf{Y}} c d\pi \le \limsup_{k\to\infty}\left(\int_{\mathbf{Y}}\phi_k d\nu - \int_{\mathbf{X}}\psi_k d\mu\right) \le \inf_{\hat\pi\in\Pi(\mu,\nu)}\int_{\mathbf{X}\times\mathbf{Y}} c d\hat\pi.$$

Hence, there is duality. □

Remark. *Theorem 1.1.9 can be generalised to lower semi-continuous functions* $c : \mathbf{X}\times\mathbf{Y}\to\mathbb{R}$*, such that, for each* $(x,y)\in\mathbf{X}\times\mathbf{Y}$, $c(x,y)\ge a(x)+b(y)$*, for some upper semi-continuous functions* $a\in L^1(\mu)$ *and* $b\in L^1(\nu)$*. Just by applying the Theorem 1.1.9 to the nonnegative lower semi-continuous function* $\hat c := c-a-b$*. Moreover if* c *is bounded from above and the optimal total cost is finite, then the Monge Kantorovich problem is solvable (there is strong duality). See [49, Section 5, Theorem 5.10] for more properties about the Monge Kantorovich problem.*

As an almost straightforward application of the Monge Kantorovich duality, the next theorem follows.

Theorem 1.1.10 (Stability)**.** *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be Polish spaces, and let* c *be a continuous cost function, with* $\inf c > -\infty$*. Let* $\{c_k\}_{k=1}^{\infty}$ *be a sequence of continuous function that uniformly converge to* c*. Let* $\{\mu_k\}_{k=1}^{\infty}$*, and* $\{\nu_k\}_{k=1}^{\infty}$ *be a sequence of probability measures on* $\mathbf{X}$ *and* $\mathbf{Y}$*, respectively. Assume that* $\mu_k \rightharpoonup \mu$ *and* $\nu_k \rightharpoonup \nu$*.*

For each $k \in \mathbb{N}$, let π_k be an optimal transport transference plan between μ_k and ν_k. If

$$(\forall k \in \mathbb{N}) \qquad \int_{\mathbf{X}\times\mathbf{Y}} c_k d\pi_k < \infty,$$

then, up to the extraction of a sub-sequence, $\{\pi_k\}_{k=1}^{\infty}$ converges to some c-cyclically monotone transference plan $\pi \in \Pi(\mu,\nu)$. Moreover, if

$$\liminf_{k\in\mathbb{N}} \int_{\mathbf{X}\times\mathbf{Y}} c_k d\pi_k < \infty,$$

then the optimal total transport $C(\mu,\nu)$ between ν and μ is finite, and π is an optimal transference plan.

For the proof we refer the reader to [49, Chapter 5, Theorem 5.20].

1.2 Wasserstein distances

In many problems it is important to know whether or not it exits a metric that induces a fixed topology for the given space. The space of finite Borel measures on a Polish space is not the exception; we wish to find a metric that induces the weak topology in $\mathscr{P}(\mathbf{X})$. There are several metrics satisfying this condition. See [6, Section 6]. Here, to solve this problem we are interested in the next approach:

Definition 1.2.1 (Wasserstein distances). *Let $(\mathbf{X}, d)$ be a Polish space, and let $p \in [1,\infty)$. For any two probability measures μ and ν, the Wasserstein metric of order p between μ and ν is defined by the formula*

$$\begin{aligned}\mathcal{W}_p(\mu,\nu) &:= \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbf{X}^2} d(x,y)^p d\pi(x,y)\right)^{\frac{1}{p}} \\ &= \inf\left\{[\mathbb{E}d(X,Y)^p]^{\frac{1}{p}};\ \ law(X)=\mu,\ \ law(Y)=\nu\right\}.\end{aligned}$$

When $d = 1$, $\mathcal{W}_1$ is also commonly called the Kantorovich-Rubinstein distance.

For the proof of next Lemma see [49, Section 1]

Lemma 1.2.2 (Gluing). *Let $(\mathbf{X},\mu)$, $(\mathbf{Y},\nu)$ and $(\mathbf{Z},\gamma)$ be Polish probability spaces. If (X,Y) is a coupling of (μ,ν) and (Y,Z) is a coupling of (ν,γ). Then there exists a triple of random variables $(\hat{X},\hat{Y},\hat{Z})$ such that $(\hat{X},\hat{Y})$ has the same law as (X,Y) and $(\hat{Y},\hat{Z})$ has the same law as (Y,Z).*

Theorem 1.2.3. *If $\mathbf{X}$ is a Polish space The Wasserstein distance is a metric over $\mathscr{P}(\mathbf{X})$.*

Proof. Let us take μ and ν in $\mathscr{P}(\mathbf{X})$. Then

$$\begin{aligned}\mathcal{W}_p(\mu,\nu) &= \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbf{X}^2} d(x,y)^p d\pi(x,y)\right)^{\frac{1}{p}} \\ &= \left(\inf_{\pi\in\Pi(\nu,\mu)} \int_{\mathbf{X}^2} d(y,x)^p d\pi(y,x)\right)^{\frac{1}{p}} = \mathcal{W}_p(\nu,\mu).\end{aligned}$$

Now take μ_1, μ_2 and μ_3 in $\mathscr{P}(\mathbf{X})$.Then by Theorem 1.1.9 there exist an optimal coupling (X_1, X_2) for (μ_1, μ_2) and an optimal coupling (X_2, X_3) for (μ_2, μ_3). By Lemma 1.2.2 there exist the random vector (Z_1, Z_2, Z_3) with law $(Z_1, Z_2) =$ law (X_1, X_2) and law $(Z_2, Z_3) =$ law (X_2, X_3), then law $(Z_1) = \mu_1$ and law $(Z_3) = \mu_3$. Hence (Z_1, Z_3) is a coupling of (μ_1, μ_3) and

$$\mathcal{W}_p(\mu_1, \mu_3) \leq \mathbb{E}\left[d(Z_1, Z_3)^p\right]^{\frac{1}{p}}$$

using Minkowski's inequality for $L^p(\mathbb{P})$ spaces;

$$\mathbb{E}\left[d(Z_1, Z_3)^p\right]^{\frac{1}{p}} \leq \mathbb{E}\left[d(Z_1, Z_2) + d(Z_2, Z_3)^p\right]^{\frac{1}{p}} \leq \mathbb{E}\left[d(Z_1, Z_2)^p\right]^{\frac{1}{p}} + \mathbb{E}\left[d(Z_2, Z_3)^p\right]^{\frac{1}{p}}$$

$$= \mathbb{E}\left[d(X_1, X_2)^p\right]^{\frac{1}{p}} + \mathbb{E}\left[d(X_2, X_3)^p\right]^{\frac{1}{p}} = \mathcal{W}_p(\mu_1, \mu_2) + \mathcal{W}_p(\mu_2, \mu_3).$$

Finally assume that $\mathcal{W}_p(\mu, \nu) = 0$ for μ and ν in $\mathscr{P}(\mathbf{X})$. Then for some $\pi \in \Pi(\mu, \nu)$ such that $d(x, y) = 0$ if and only if $x = y$ π-almost surely, then there is at least one transfer plan that it is concentrated at the set $\{(x, x), x \in \mathbf{X}\} \subset \mathbf{X}^2$, so $\nu = \mathrm{Id}_{\#}\mu = \mu$. Conversely $\mathcal{W}_p(\mu, \mu) = 0$, just by taking the product measure $\mu \otimes \mu$. □

Definition 1.2.4. *Let (X, d) be a Polish space, and let $p \in [1, \infty)$. We define the Wasserstein space of order p as*

$$P_p(\mathbf{X}) := \left\{ \mu \in \mathscr{P}(\mathbf{X}); \quad \int_{\mathbf{X}} d(x_0, x)^p \mu(x) < \infty \right\},$$

here, $x_0 \in \mathbf{X}$ is arbitrary.

Remark. *It can be proved that P_p does not depend on x_0.*

Theorem 1.2.5. *Let (X, d) be a Polish space, and let $p \in [1, \infty)$. Then for any μ and ν in $P_p(\mathbf{X})$, $\mathcal{W}_p(\mu, \nu) < \infty$.*

Proof. Let us take $\pi \in \Pi(\mu, \nu)$ then using that

$$d(x, y)^p \leq 2^{p-1}[d(x, x_0)^p + d(x_0, y)^p],$$

the result follows by taking the integral with respect to π. □

Now that we have a metric, what can we say about the topology in $P_p(\mathbf{X})$?

Definition 1.2.6 (Weak convergence in $P_p(\mathbf{X})$)**.** *Let $(\mathbf{X}, d)$ be a Polish space, and $p \in [1, \infty)$. Let $\{\mu_k\}_{k=1}^{\infty}$ be a sequence of probability measures in $P_p(\mathbf{X})$, and $\mu \in P_p(\mathbf{X})$. Then $\{\mu_k\}_{k=1}^{\infty}$ is said to converge weakly to μ in $P_p(\mathbf{X})$ if any of the following equivalent properties is satisfied for some- and then any - $x_0 \in \mathbf{X}$:*

i) $\mu_k \rightharpoonup \mu$ *and* $\displaystyle\int_{\mathbf{X}} d(x_0, x)^p d\mu_k(x) \to \int_{\mathbf{X}} d(x_0, x)^p d\mu(x)$;

ii) $\mu_k \rightharpoonup \mu$ *and* $\displaystyle\limsup_{k\to\infty} \int_{\mathbf{X}} d(x_0, x)^p d\mu_k(x) \leq \int_{\mathbf{X}} d(x_0, x)^p d\mu(x)$;

iii) $\mu_k \rightharpoonup \mu$ *and* $\displaystyle\lim_{R\to\infty} \limsup_{k\to\infty} \int_{\{x\in\mathbf{X}; d(x,x_0)\geq R\}} d(x_0, x)^p d\mu_k(x) = 0$;

iv) For all continuous functions ψ with $|\psi(x)| \leq C(1 + d(x_0, x)^p)$, $C \in \mathbb{R}$, one has

$$\int_{\mathbf{X}} \psi(x) d\mu_k(x) \to \int_{\mathbf{X}} \psi(x) d\mu(x).$$

Now we state a very useful theorem –by which we refer for the proof to [49, Chapter 6,Theorem 6.9]–, with two straightforward implications.

Theorem 1.2.7 ($\mathcal{W}_p$ metrizes P_p)**.** *Let $(\mathbf{X}, d)$ be a Polish space, and $p \in [1, \infty)$. Then the Wasserstein distance $\mathcal{W}_p$ metrizes the weak convergence in $P_p(\mathbf{X})$,i.e., if $\{\mu_k\}_{k=1}^{\infty}$ is a sequence in $P_p(\mathbf{X})$ and μ in $\mathscr{P}(\mathbf{X})$ then $\{\mu_k\}_{k=1}^{\infty}$ converges weakly to μ in $P_p(\mathbf{X})$ if and only in $\mathcal{W}_p(\mu, \mu_k) \to 0$.*

Corollary 1.2.8 (Continuity of $\mathcal{W}_p$)**.** *Let $(\mathbf{X}, d)$ be a Polish space, and $p \in [1, \infty)$; then the Wasserstein distance $\mathcal{W}_p$ is continuous in $P_p(\mathbf{X})$.*

Corollary 1.2.9 (Metrizability of the weak topology)**.** *Let $(\mathbf{X}, d)$ be a Polish space, and $p \in [0, \infty)$. If $\hat{d}$ is a bounded distance inducing the same topology as d (i.e., $\hat{d} = \frac{d}{d+1}$), then the convergence in the Wasserstein sense for the distance $\hat{d}$ is equivalent to the weak convergence of probability measures in $\mathscr{P}(\mathbf{X})$.*

Finally we conclude this chapter with a topological property of the Wassertein space; for the proof we refer to [49, Chapter 6,Theorem 6.18]

Theorem 1.2.10 (Topology of the Wassertein space)**.** *Let $\mathbf{X}$ be a Polish space and $p \in [1, \infty)$. Then $(P_p(\mathbf{X}), \mathcal{W}_p)$ is also a Polish space. Moreover, any probability measure can be approximated by a sequence of probability measures with finite support.*

1.3 Displacement Interpolation

In this section we introduce a time-dependent version of optimal transport. This adaptation will help us to describe the dynamics in some complex systems, and will link other areas of mathematics with optimal transport in a very straightforward way.

We shall assume that the initial and final probability measures are defined on the same Polish space $(\mathbf{X}, d)$. The main additional structure is that the cost is associated with an action, i.e., the cost function between an initial point $x \in \mathbf{X}$ and a final point $y \in \mathbf{X}$ is obtained by minimising the action among paths that go from x to y:

$$c(x, y) := \inf\{\mathcal{A}(\gamma); \quad \gamma_0 = x, \quad \gamma_1 = y \quad ; \gamma \in \mathcal{C}\}. \tag{1.3.1}$$

Here $\mathcal{C}$ is a certain class of continuous curves on which the action functional $\mathcal{A}$ is defined. To give an idea of how these action functionals $\mathcal{A}$ would look like, imagine a smooth surface. Then, given x and y over the surface, $c(x, y)$ could be the infimum of the length of smooth curves with ending points x and y. It will be useful to consider an action as a family of functionals parametrized by the initial and the final times: so $\mathcal{A}^{s,t}(\gamma)$ is a functional on the set of continuous paths $[s,t] \to \mathbf{X}$. Then we let

$$c^{s,t}(x,y) = \inf\left\{\mathcal{A}^{s,t}(\gamma); \quad \gamma_s = x, \quad \gamma_t = y \quad ; \gamma \in \mathbf{C}([s,t];\mathbf{X})\right\}. \tag{1.3.2}$$

In words, $c^{s,t}$ is the minimal work needed to complete the action from the point x at the initial time s, to the point y at the final time t.

For simplicity of notation, throughout this section the curve γ will stand for a curve $\gamma \in \mathbf{C}([s,t];\mathbf{X})$.

Definition 1.3.1 (Abstract Lagrangian actions)**.** *Let $(\mathbf{X}, d)$ be a Polish space, ant let $t_i, t_f \in \mathbb{R}$. A Lagrangian action $(\mathcal{A})^{t_i,t_f}$ on $\mathbf{X}$ is a family of lower semi-continuous functionals $\mathcal{A}^{s,t}$ on $\mathbf{C}([s,t],\mathbf{X})$ with $(t_i \le s < t \le t_f)$, with respect to the supremum norm and cost functions $c^{s,t}$ on $\mathbf{X}\times\mathbf{X}$, such that:*

i) $t_i \le t_1 < t_2 < t_3 \le t_f \implies \mathcal{A}^{t_1,t_2} + \mathcal{A}^{t_2,t_3} = \mathcal{A}^{t_1,t_3}$;

ii) $(\forall x, y \in \mathbf{X})$

$$c^{s,t}(x,y) = \inf\left\{\mathcal{A}^{s,t}(\gamma); \quad \gamma_s = x, \quad \gamma_t = y \quad ; \gamma \in \mathbf{C}([s,t];\mathbf{X})\right\}.$$

iii) For any curve $(\gamma_t)_{t_i\le t\le t_f}$

$$\mathcal{A}^{t_i,t_f}(\gamma) = \sup_{N\in\mathbb{N}} \sup_{t_i=t_0\le t_1\le\cdots\le t_N=t_f} \left\{\sum_{k=0}^{N-1} c^{t_k,t_{k+1}}(\gamma_{t_k},\gamma_{t_{k+1}})\right\}.$$

This kind of structures frequently arises in physics and optimal control at the time of minimising certain cost functions or energies over some dynamical system. This structure is also related to Hamiltonian equations. For examples and for more detailed information we refer to [35, 4].

Definition 1.3.2 (Coercive actions)**.** *Let $(\mathcal{A})^{0,1}$ be a Lagrangian action on a Polish space $\mathbf{X}$, with associated cost functions $(c^{s,t})_{0\le s<t\le 1}$. For any two times s, t $(0 \le s < t \le 1)$, and any two compact sets $K_s, K_t \subset \mathbf{X}$, let $\Gamma^{s,t}_{K_s\to K_t}$ be the set of minimising paths starting in K_s at time s, and ending in K_t at time t. The action will be called coercive if:*

i) Is bounded below, in the sense that

$$\inf_{0\le s<t\le 1} \inf_{\gamma\in\mathbf{C}([s,t],\mathbf{X})} \mathcal{A}^{s,t}(\gamma) > -\infty;$$

ii) If $s < t$ are any two intermediate times, and K_s, K_t are any two compact sets such that $c^{s,t}(x,y) < \infty$ for all $x \in K_s$ and $y \in K_t$, then the set $\Gamma^{s,t}_{K_s\to K_t}$ is compact and nonempty. In particular, minimising curves between any two fixed points x, $y \in \mathbf{X}$, with $c^{0,1}(x,y) < \infty$ should always exist and form a compact set.

Proposition 1.3.3. *Let $(\mathbf{X}, d)$ be a Polish space and $(\mathcal{A})^{0,1}$ a coercive Lagrangian action on $\mathbf{X}$. Then:*

i) For all intermediate times $s < t$, $c^{s,t}$ is lower semi-continuous on $\mathbf{X} \times \mathbf{X}$, with values in $\mathbb{R} \cup \{\infty\}$.

ii) If a curve γ on $[s,t] \subset [0,1]$ is a minimiser of $\mathcal{A}^{s,t}$, then its restriction to $[s',t'] \subset [s,t]$ is also a minimiser of $\mathcal{A}^{s',t'}$.

iii) For all times $t_1 < t_2 < t_3$ in $[0,1]$, and $x_1, x_3 \in \mathbf{X}$,

$$c^{t_1,t_3}(x_1,x_3) = \inf_{x_2 \in \mathbf{X}} \{c^{t_1,t_2}(x_1,x_2) + c^{t_2,t_3}(x_2,x_3)\};$$

and if the infimum is achieved at some point x_2, then there is a minimising curve which goes from x_1 at time t_1 to x_3 at time t_3, and passes through x_2 at time t_2.

iv) A curve γ is a minimiser of $\mathcal{A}^{0,1}$ if and only if, for intermediate times $t_1 < t_2 < t_3$ in $[0,1]$,

$$c^{t_1,t_3}(\gamma_{t_1},\gamma_{t_3}) = c^{t_1,t_2}(\gamma_{t_1},\gamma_{t_2}) + c^{t_2,t_3}(\gamma_{t_2},\gamma_{t_3}).$$

v) If the cost functions $c^{s,t}$ are continuous, then the set Γ of all action-minimising curves is closed in the topology of uniform convergence.

vi) For all times $s < t$, there is a Borel map $S_{s\to t} : \mathbf{X} \times \mathbf{X} \to \mathbf{C}([s,t],\mathbf{X})$, such that for all $x, y \in \mathbf{X}$, $S_{s\to t}(x,y)$ belongs to $\Gamma^{s,t}_{x\to y}$.

Proof. i) Let us take $\{x_k\}_{k=1}^{\infty}$ and $\{y_k\}_{k=1}^{\infty}$, sub-sequences of $\mathbf{X}$ converging to x and y, respectively. Then $K_s := \{x_k\}_{k=1}^{\infty} \cup \{x\}$ and $K_t := \{y_k\}_{k=1}^{\infty} \cup \{y\}$ are compact. By hypothesis there is a minimising curve $\gamma_k \in \mathbf{C}([s,t],\mathbf{X})$ joining x_k to y_k for each $k \in \mathbb{N}$, therefore $\gamma_k \in \Gamma^{s,t}_{K_s \to K_t}$ which is also compact. Then there is a sub-sequence of $\{\gamma_k\}_{k=1}^{\infty}$ converging uniformly to a minimising curve γ that joins x to y. Thus lower semi-continuity of $\mathcal{A}^{s,t}$ implies

$$c^{s,t}(x,y) \le \mathcal{A}^{s,t}(\gamma) \le \liminf_{k\to\infty} \mathcal{A}^{s,t}(\gamma_k) \le \mathcal{A}^{s,t}(\gamma) \le \liminf_{k\to\infty} c^{s,t}(x_k,y_k).$$

ii) Let us take a minimising curve γ for $\mathcal{A}^{s,t}$ given the fixed points x and y. Denote $x' := \gamma_{s'}$ and $y' := \gamma_{t'}$. If the restriction is not optimal for $\mathcal{A}^{s',t'}$, then there is an optimal curve γ' joining x' and y'. Therefore $\mathcal{A}^{s',t'}(\gamma') < \mathcal{A}^{s',t'}(\gamma)$. Hence the curve obtained by the restriction of γ in $[s,s']$, γ' in $[s',t']$ and γ in $[t',t]$ has a strictly lower action in $\mathcal{A}^{s,t}$ than γ, which is impossible.

iii) Let us take an arbitrary point $x \in \mathbf{X}$. Denote γ_1 the minimising curve joining x_1 and x_2, at times t_1 and t_2, and γ_2 the minimising curve joining x_2 and x_3, at time t_2 and t_3. Set γ_x by concatenating the curves γ_1 and γ_2, in the point x_2 at time t_2. Then by definition 1.3.1 of Lagrangian actions,

$$c^{t_{1,3}}(x_1,x_3) \le \mathcal{A}^{t_1,t_3}(\gamma_x) = \mathcal{A}^{t_1,t_2}(\gamma_1) + \mathcal{A}^{t_2,t_3}(\gamma_2) = c^{t_{1,2}}(x_1,x_2) + c^{t_{2,3}}(x_2,x_3).$$

The equality follows by taking the infimum over all $x_2 \in \mathbf{X}$ and using the property *ii)* and splitting the minimising curve for $c^{t_1,t_3}(x_1,x_3)$ at point t_2. Moreover, if there is equality at some point $x_2 \in \mathbf{X}$ then γ_x is the minimising curve and it passes through x_2 at time t_2.

iv) Since property *iii)* holds true and any restriction to the sets $[t_1, t_2]$ and $[t_2, t_3]$ is minimising, it is easily seen that any minimising curve should satisfy property *iv)*. Conversely, let γ be a curve satisfying property *iv)*. We now proceed by induction, that implies that for any subdivision $0 = t_0 < t_1 < \cdots < t_N = 1$, of the interval $[0, 1]$

$$c^{0,1}(\gamma_0, \gamma_1) = \sum_{k=0}^{n} c^{t_k, t_{k=1}}(\gamma_{t_k}, \gamma_{t_{k+1}}).$$

By property *iii)* in the definition 1.3.1 of Lagrangian actions, γ minimises $\mathcal{A}$.

v) Let us take $0 \leq t_1 < t_2 < t_3 \leq 1$, and denote by $\Gamma(t_1, t_2, t_3)$ the set of curves satisfying property *iv)*. Then, as all functions $c^{s,t}$ are continuous and uniform convergence implies point-wise convergence, $\Gamma(t_1, t_2, t_3)$ is closed for any fixed times t_1, t_2 and t_3. Then $\Gamma = \cap_{0 \leq t_1 < t_2 < t_3 \leq 1} \Gamma(t_1, t_2, t_3)$ is closed.

vi) Given the fixed times $s < t$, denote by $\Gamma^{s,t}$ the set of all action-minimising curves defined over $[s, t]$, and let $E_{s,t} : \Gamma \to \mathbf{X} \times \mathbf{X}$ be the end-point map defined as $E_{s,t}(\gamma) := (\gamma_s, \gamma_t)$. By hypothesis $E_{s,t}$ is surjective. Then, by definition of coercive actions, the pre-image of a compact set in $\mathbf{X} \times \mathbf{X}$ is compact. Therefore $E_{s,t}$ is a continuous function between Polish spaces. In particular, for each x and y in $\mathbf{X}$, $E_{s,t}^{-1}(x, y)$ is compact. By [19, Theorem 40], there exists a measurable right-inverse function $S_{s \to t}$ joining x and y with a minimising curve.

□

Let c be the cost associated with the Lagrangian action, and let μ_0, $\mu_1 \in \mathscr{P}(\mathbf{X})$. Introduce an optimal coupling (X_0, X_1) of (μ_0, μ_1), and a random action-minimizing path $\{X_t\}_{0 \leq t \leq 1}$ joining X_0 to X_1. Then we sat that $\{X_t\}_{0 \leq t \leq 1}$ is an interpolation of μ_0 and μ_1. This procedure is called displacement interpolation.

Definition 1.3.4 (Dynamical Coupling)**.** *Let $(\mathbf{X}, d)$ be a Polish space. A dynamical transference plan Π is a probability measure on the space $\mathbf{C}([0, 1], \mathbf{X})$. A dynamical coupling of two probability measures μ_0, $\mu_1 \in \mathscr{P}(\mathbf{X})$ is a random curve $\gamma : [0, 1] \to \mathbf{X}$ such that law$(\gamma_0) = \mu_0$ and law$(\gamma_1) = \mu_1$*

Definition 1.3.5 (Optimal Dynamical Coupling)**.** *Let $(\mathbf{X}, d)$ be a Polish space, $(\mathcal{A})^{0,1}$ a Lagrangian action on $\mathbf{X}$, c the associated costs, and Γ the set of action-minimising curves. A dynamical optimal transport transference plan is a probability measure Π on Γ such that*

$$\pi_{0,1} := (e_0, e_1)_{\#}\Pi$$

is an optimal transference plan between μ_0 and μ_1. Equivalently, Π is the law of a random action-minimizing curve whose endpoints constitute and optimal coupling of μ_0 and μ_1, where e_t will stand for the evaluation functional at time t: $e_t(\gamma) = \gamma_t$.

For the proof of the following theorem we refer to [49, *Theorem* 7.21].

Theorem 1.3.6 (Displacement Interpolation)**.** *Let $(\mathbf{X}, d)$ be a Polish space, $(\mathcal{A})^{0,1}$ a Lagrangian action on $\mathbf{X}$, with continuous cost functions $c^{s,t}$. Whenever $0 \leq s < t \leq 1$, denote $C^{s,t}(\mu, \nu)$ the optimal transport cost between μ and ν with cost function $c^{s,t}$; write $c := c^{0,1}$ and $C = C^{0,1}$. Let μ_0, $\mu_1 \in \mathscr{P}(\mathbf{X})$ be such that the optimal* *transport cost $C(\mu_0, \mu_1)$ is finite. Then, given a continuous path $\{\mu_t\}_{0\le t\le 1}$, the following properties are equivalent:*

i) For each $t \in [0,1]$, $\mu_t = law(\gamma_t)$, where $(\gamma_t)_{0\le t\le 1}$ is a dynamical optimal coupling for (μ_0, μ_1)

ii) For any three intermediate times $0 \le t_1 < t_2 < t_3 \le 1$

$$C^{t_1,t_2}(\mu_{t_1}, \mu_{t_2}) + C^{t_2,t_3}(\mu_{t_2}, \mu_{t_3}) = C^{t_1,t_3}(\mu_{t_3}, \mu_{t_3});$$

iii) The path $(\mu_t)_{0\le t\le 1}$ is a minimizing curve for the coercive action functional defined on $\mathscr{P}(\mathbf{X})$ by

$$\mathbb{A}^{s,t}(\mu) = \sup_{N\in\mathbb{N}} \sup_{t_i=t_0\le t_1\le\cdots\le t_N=t_f} \left\{ \sum_{k=0}^{N-1} C^{t_k,t_{k+1}}(\mu_{t_k}, \mu_{t_{k+1}}) \right\} \tag{1.3.3}$$

$$= \inf_{\gamma} \mathbb{E}(\mathcal{A}^{s,t}(\gamma)), \tag{1.3.4}$$

where the last infimum is over all random curves $\gamma : [s,t] \to \mathbf{X}$ such that the $law(\gamma_\tau) = \mu_\tau$ $(0 \le \tau \le 1)$. In that case we say that $(\mu_t)_{0\le t\le 1}$ is a displacement interpolation between μ_0 and μ_1. There always exist such curve. Finally, if $\mathcal{K}_0$ and $\mathcal{K}_1$ are two compact subsets of $\mathscr{P}(\mathbf{X})$, such that $C(\mu_0, \mu_1) < \infty$ for all $\mu_0 \in \mathcal{K}_0$, $\mu_1 \in \mathcal{K}_1$, then the set of dynamical optimal transport transference plans Π with $(e_0)_\#\Pi \in \mathcal{K}_0$ and $(e_1)_\#\Pi \in \mathcal{K}_1$ is compact.

Theorem 1.3.6 admits an important corollary:

Corollary 1.3.7. *With the same assumptions as in Theorem 1.3.6, suppose that*

a) there is a unique optimal transport plan π between μ_0 and μ_1,

b) $\pi-$almost surely, x_0 and x_1 are joined by a unique minimizing curve.

Then there exists a unique displacement interpolation $(\mu_t)_{0\le t\le 1}$ joining μ_0 and μ_1.

Concluding remarks

In this chapter, we have seen the big descriptive picture of optimal transport. We gave sufficient conditions for the solvability to the Monge Kantorovich and the dual Monge Kantorovich problems. We explored the time-depended version of optimal transport problem. And we endowed the space of Borel probability measures of a Polish space with an optimal transport based metric, which provided a reach topological structure to the space of Borel probability measures.

Optimal transport encompasses several areas of application. It is important to remark that the relevance of its results relies on the generality under the assumptions for the solvability of the problem. However, this generality in the results leads to inefficiency of explicitly calculating those optimal transference plans. Just consider measures with finite support, as in the proof of the duality Theorem, then the optimal transference plan can be explicitly calculated by finite linear programming methods. This is not the case when we take into consideration arbitrary Borel measures. Thus, optimal transport has improved new results in different areas, for example, partial differential equations.

Chapter 2

Optimal Transport and Game Theory

A time-honoured approach to study non-cooperative games is via the concept of Nash equilibrium. Thus, the main objective of this chapter is to briefly introduce the basic concepts of game theory and its relations with optimal transport.
On the one hand, we introduce the concept of an $N-$person game which is conformed of a set of N players, a family of strategy sets, and a set of cost functions. We formally define a Nash Equilibrium –see Definition 2.1.1– for an N-person game, and –under suitable assumptions over the elements that conform the game– we state the Nash equilibria fundamental theorem. On the other hand, we set a model for games with a continuum of players –also known as non-atomic games–, and we extrapolated the notion of Nash equilibrium to these games by exploring the limit of Nash equilibria when the number of players tends to infinity (see Theorems 2.3.1 and 2.3.2).
Finally, we characterise Cournot Nash equilibria-see Definition 2.2.1— for an specific family of non-atomic games with an unique optimal transport plan (see Lemma 2.2.8).

2.1 Nash Equilibria

Let us consider a set of N players $i \in \{1, \dots, N\}$. Each player i has a strategy space X_i. and a cost function $J_i : \mathbf{X} \to \mathbb{R}$, where $\mathbf{X} := \Pi_{i=1}^N X_i$. Then, the triple $(\{1, \dots, N\}, \mathbf{X}, \{J_i\}_{i=1}^N)$ is known as an N-person game.

Remark. *$x = (x_1, \dots, x_N) \in \mathbf{X}$, is also expressed as $x = (x_i|x_{-i}) \in \mathbf{X}$, where $x_{-i} = (x_1, \dots, x_{j-1}, x_{j+1} \dots, j_N) \in X_{-i} := \Pi_{j \neq i} X_j$.*

A Nash equilibrium is defined as follow:

Definition 2.1.1 (Nash equilibrium). *A Nash equilibrium is an element $\bar{\mathbf{x}} := (\bar{x}_1, \dots, \bar{x}_N) = (\bar{x}_i|\bar{x}_{-i}) \in \mathbf{X}$ such that, for every $i \in \{1, \dots, N\}$*

$$J_i(\bar{x}_i, \bar{x}_{-i}) \leq J_i(x_i, \bar{x}_{-i}),$$

for each $x_i \in X_i$.

Definition 2.1.2 (Quasi-convex)**.** *A real valued function f over a convex space S is quasi-convex if for all x, $y \in S$ and $\lambda \in [0,1]$ we have*

$$f(\lambda x + (1-\lambda)y) \leq \max(f(x), f(y)).$$

The fundamental theorem of Nash states the following existence result for Nash equilibria:

Theorem 2.1.3 (Existence of Nash equilibria)**.** *Let X_i be a convex compact subset of some locally convex Hausdorff topological vector space. Let J_i be continuous on $\mathbf{X}$ for every $i \in \{1, \ldots, N\}$. And let $J(\cdot, x_{-i})$ be a quasi-convex function on X_i for every $x_{-i} \in X_{-i}$. Then, there exist at least one Nash equilibrium.*

Because the restriction about convexity in the above, Theorem 2.1.3 excludes important cases of study, for example, when the spaces of strategies are finite; Nash also introduced the concept of mixed strategy:

Definition 2.1.4 (Mixed strategies)**.** *A mixed strategy of player i is a probability measure $\pi_i \in \mathscr{P}(X_i)$. Given a mixed strategy $N-$vector $\pi = (\pi_1, \ldots, \pi_N) \in \Pi_{n=1}^N \mathscr{P}(X_i)$, the cost of player i associated to π is given by*

$$\bar{J}_i(\pi_1, \ldots, \pi_N) := \int_{\mathbf{X}} J_i(x_1, \ldots, x_N) \otimes_{n=1}^N \pi_j(dx_j).$$

Remark. *It is easy to check that a convex combination of mixed strategies, remains a mixed strategy. Therefore, the sets of mixed strategies is convex.*

Theorem 2.1.5. *A game with compact metric strategy spaces and continuous cost functions has at least one Nash equilibrium in mixed strategies. In particular, finite games admit Nash equilibria in mixed strategies.*

2.2 Cournot-Nash and Optimal Transport

Our aim is to analyse a system with a huge amount of agents, each of which is characterized by a type x (for example, the socio-economical status, income, etc.) belonging to some compact metric space $\mathbf{X}$. This space is endowed with a Borel probability measure $\mu \in \mathscr{P}(\mathbf{X})$, which gives the distribution of the population over the types. Each agent takes an strategy y from the strategy space $\mathbf{Y}$, which is also a compact metric space. For each agent, his cost will not depend only on his type or the strategies he takes, it will also depend the strategies taken by the other agents, which is encoded by a probability distribution $\nu \in \mathscr{P}(\mathbf{Y})$. More formally the cost is given by some function $F \in C(\mathbf{X} \times \mathbf{Y} \times \mathscr{P}(\mathbf{Y}))$, where $\mathscr{P}(\mathbf{Y})$ is endowed with the weak-$*$ topology. Here the notion of an equilibrium is given by a joint probability measure on $\mathbf{X} \times \mathbf{Y}$, whose marginals are consistent with the type distribution and the cost-minimising behaviour of the agents. This leads to the following definition:

Definition 2.2.1 (Cournot-Nash equilibrium)**.** *A Cournot-Nash equilibrium for F and μ is a $\gamma \in \mathscr{P}(\mathbf{X} \times \mathbf{Y})$ such that the first marginal of γ is a probability measure $\Pi_{\mathbf{X}\#}\gamma = \mu$ and γ gives the full mass to cost-minimising strategies:*

$$\gamma\left(\left\{(x,y) \in \mathbf{X} \times \mathbf{Y} : F(x,y,\nu) = \min_{y' \in \mathbf{Y}} F(x,y',\nu)\right\}\right) = 1,$$

where ν denotes the second marginal of γ: $\Pi_{\mathbf{X}\#}\gamma = \nu$.

If γ is Cournot-Nash Equilibrium for F, and $\nu := \Pi_{\mathbf{Y}\#}\gamma$ is of the form $\nu = T_{\#}\mu$, for some measurable function $T : \mathbf{X} \to \mathbf{Y}$, we say that the γ is a *pure* Cournot-Nash equilibrium. This also can be interpreted as: agents with the same type use the same strategy.

Theorem 2.2.2 (Existence of Cournot-Nash equilibria). *If $F \in C(\mathbf{X}\times\mathbf{Y}\times\mathscr{P}(\mathbf{Y}))$, where $\mathscr{P}(\mathbf{Y})$ is endowed with the weak-$*$ topology, then there exists at least one Cournot-Nash equilibrium.*

An application of optimal transport arises in the calculation of Cournot-Nash Equilibria, when we take into consideration a particular class of functions F in Definition 2.2.1. Taking the assumptions of the above problems, i.e., model that consist of a compact metric type space $\mathbf{X}$, with a Borel probability measure $\mu \in \mathscr{P}(\mathbf{X})$, a compact metric action space $\mathbf{Y}$, and also here will be needed a non-negative reference measure μ_0. In the sequel, we will restrict ourselves to the case where $F(x, y, \nu) = c(x, y) + \mathcal{V}[\nu](y)$, where c and $\mathcal{V}$ are continuous. We set a social cost associated to γ by

$$\begin{aligned} \mathrm{SC}[\gamma] :&= \int_{\mathbf{X}\times\mathbf{Y}} F(x, y, \nu) d\gamma(x, y) \\ &= \int_{\mathbf{X}\times\mathbf{Y}} c(x, y) d\gamma(x, y) + \int_{\mathbf{Y}} \mathcal{V}[\nu](y) d\nu(y), \end{aligned}$$

where $\nu = \Pi_{\mathbf{Y}\#}\gamma$.
As c is continuous over a compact space, it is bounded, and so $\int_{\mathbf{X}\times\mathbf{Y}} c(x, y)d\gamma(x, y)$ is finite, for every $\gamma \in \mathscr{P}(\mathbf{X} \times \mathbf{Y})$. However, for $\mathrm{SC}[\gamma]$ to be finite, we require that second marginal ν –also known as action marginal – belongs to the set

$$\mathcal{D} := \left\{ \nu \in \mathscr{P}(\mathbf{Y}) : \ \ \nu \ll \mu_0, \ \ \int_{\mathbf{Y}} |\mathcal{V}[\nu]| d\nu < \infty \right\}. \tag{2.2.1}$$

Let us define $\phi(x) := \min_{y' \in \mathbf{Y}} F(x, y', \nu)$, for some $\nu \in \mathcal{D}$, then if γ is a Cournot-Nash equilibrium for $F(x, y, \nu) = c(x, y)+\mathcal{V}[\nu](y)$, $\phi(x) \leq c(x, y)+\mathcal{V}[\nu](y)$ $\nu-$almost every y and $\phi(x) = c(x, y)+\mathcal{V}[\nu](y)$ $\gamma-$almost every (x, y). Taking into consideration the reference measure μ_0, we state another Cournot-Nash definition:

Definition 2.2.3. *$\gamma \in \mathscr{P}(\mathbf{X}) \times \mathscr{P}(\mathbf{Y})$ is a μ_0-Cournot-Nash equilibria, if its first marginal is μ, its second marginal, ν, is in $\mathcal{D}$, and there exist $\phi \in \mathcal{C}(\mathbf{X})$ such that*

$$c(x, y) + \mathcal{V}[\nu](y) \geq \phi(x) \quad \textit{for all } x \in \mathbf{X}, \textit{ and } \mu_0-\textit{a.e- } y \textit{ with equality } \gamma-\textit{a.e.}.$$

We also say that γ is pure whenever it is of the form $\gamma = (Id, T)_{\#}\mu$ for some measurable map $T : \mathbf{X} \to \mathbf{Y}$.

Suppose that we have a Cournot-Nash equilibrium γ for the function F. Let ν be his marginal over (Y). Then ν must minimizes $\mathcal{V}$,furthermore, γ must be an optimal transport between μ and ν, otherwise, γ is not and Cournot-Nash Equilibria. Inspire by this straight forward observation Carlier and Blanchet [8] give the next lemma.

Lemma 2.2.4. *Let γ be a Cournot-Nash equilibrium for F, and set $\nu := \Pi_{Y\#}\gamma$. Then γ is an optimal transport between μ and ν.*

Proof. Let us take ϕ as in Definition 2.2.3, and $\eta \in \Pi(\mu, \nu)$. One the one hand, we have that

$$\int_{\mathbf{X}\times\mathbf{Y}} c(x,y)d\gamma(x,y) = \int_{\mathbf{X}} \phi(x)d\mu(x) - \int_{\mathbf{Y}} \mathcal{V}[\nu](y)d\nu(y),$$

and on the other hand, we have

$$c(x,y) \geq \phi(x) - \mathcal{V}[\nu](y)$$

for every $x \in \mathbf{X}$ and $\nu-$a.e.y. Then integrating with η one has

$$\begin{aligned}\int_{\mathbf{X}\times\mathbf{Y}} c(x,y)d\eta(x,y) &\geq \int_{\mathbf{X}\times\mathbf{Y}} (\phi(x) - \mathcal{V}[\nu](y))d\eta(x,y) \\ &= \int_{\mathbf{X}} \phi(x)d\mu(x) - \int_{\mathbf{Y}} \mathcal{V}[\nu](y)d\nu(y) \\ &= \int_{\mathbf{X}\times\mathbf{Y}} c(x,y)d\gamma(x,y).\end{aligned}$$

Hence γ is a minimizer. □

Our problem reduces to find such a ν that simultaneously minimises $\mathcal{V}[\nu]$ and the map $\nu \mapsto \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbf{X}\times\mathbf{Y}} c(x,y)d\gamma(x,y)$. Note that a ν that minimizes $\mathcal{V}[\nu]$ must not be the marginal of a Cournot-Nash equilibrium. Then –in analogy to the ideas in the seminal work of Monderer and Shapley [36] on potential games– A. Blanchet and G. Carlier [8] gave a variational approach to Cournot Nash equilibria that could be obtained by minimizing some potential function over the set of Borel probability measures. The main idea is to suppose that $\mathcal{V}$ is, in some sense, the first variation of some function $\mathcal{E}$.

Definition 2.2.5 (Blanchet-Carlier Differential)**.** *Let $\mathcal{D}$ be defined as in* (2.2.1). *We say that the map $\nu \mapsto \mathcal{V}[\nu] \in C(\mathbf{Y})$ for $\nu \in \mathcal{D}$ is a differential on $\mathcal{D}$ if $\mathcal{D}$ is convex and there exist $\mathcal{E} : \mathcal{D} \to \mathbb{R}$ such that for every ρ and ν in $\mathcal{D}$, $\mathcal{V}[\nu] \in L^1(\rho)$ and*

$$\lim_{\epsilon\to 0^+} \frac{\mathcal{E}[(1-\epsilon)\nu + \epsilon\rho] - \mathcal{E}[\nu]}{\epsilon} = \int_{\mathbf{Y}} \mathcal{V}[\nu]d(\rho - \nu).$$

In this case $\mathcal{V}[\nu]$ is a differential on $\mathcal{D}$ of $\mathcal{E}$ and we denote $\mathcal{V}[\nu] = \dfrac{\delta\mathcal{E}}{\delta\nu}$.

To clear ideas and illustrate how does $\mathcal{E}$ would look like, let us consider one example.

Example 2.2.6. *Suppose that*

$$\mathcal{V}[\nu](y) := \int_{\mathbf{Y}} \phi(y,z)d\nu(z),$$

for some function $\phi \in C(\mathbf{Y} \times \mathbf{Y})$. Then it is natural to define

$$\mathcal{E}[\nu] = \frac{1}{2} \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(y,z) d\nu(y) d\nu(z).$$

As ϕ is continuous over a compact space, it is bounded. Moreover

$$\begin{aligned}
2\mathcal{E}[(1-\epsilon)\nu + \epsilon\rho] &= \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(x,y) d[(1-\epsilon)\nu + \epsilon\rho](x) d[(1-\epsilon)\nu + \epsilon\rho](y) \\
&= \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(x,y)[d\nu(x)d\nu(y) + \epsilon^2 d\nu(x)d\nu(y) + \epsilon^2 d\rho(x)d\rho(y)] \\
&+ \epsilon \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(x,y)[d\nu(x)d\rho(y) + d\rho(x)d\nu(y) - 2d\nu(x)d\nu(y)] \\
&- \epsilon^2 \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(x,y)[d\nu(x)d\rho(y) + d\rho(x)d\nu(y)].
\end{aligned}$$

Then

$$\begin{aligned}
\frac{\mathcal{E}[(1-\epsilon)\nu + \epsilon\rho] - \mathcal{E}[\nu]}{\epsilon} &= \frac{\epsilon}{2} \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(x,y)[d\nu(x)d\nu(y) + d\rho(x)d\rho(y) - d\nu(x)d\rho(y) - d\rho(x)d\nu(y)] \\
&+ \frac{1}{2} \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(x,y)[d\nu(x)d\rho(y) + d\rho(x)d\nu(y) - 2d\nu(x)d\nu(y)].
\end{aligned}$$

Letting ϵ tend to 0^+, we have that

$$\begin{aligned}
\lim_{\epsilon \to 0^+} \frac{\mathcal{E}[(1-\epsilon)\nu + \epsilon\rho] - \mathcal{E}[\nu]}{\epsilon} &= \frac{1}{2} \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(x,y)[d\nu(x)d\rho(y) + d\rho(x)d\nu(y) - 2d\nu(x)d\nu(y)] \\
&= \frac{1}{2} \int_{\mathbf{Y}} \int_{\mathbf{Y}} \phi(x,y)[d\nu(x)d(\rho(y) - \nu(y)) + d\nu(y)d(\rho(x) - \nu(x))] \\
&= \frac{1}{2} \int_{\mathbf{Y}} \int_{\mathbf{Y}} (\phi(x,y) + \phi(y,x))[d\nu(y)d(\rho(x) - \nu(x))].
\end{aligned}$$

If ϕ is symmetric, one finally has

$$\frac{\delta\mathcal{E}}{\delta\nu}(y) = \int_{\mathbf{Y}} \phi(x,y) d\nu(x).$$

Now we state a lemma that according to which the optimal cost

$$\mathcal{W}_c(\mu,\nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbf{X} \times \mathbf{Y}} c(x,y) d\gamma(x,y)$$

has a differential on $\mathcal{D}$ in the sense of Definition 2.2.5.

Lemma 2.2.7. *Assume that $\mathbf{Y}$ is a compact metric space, $\mathbf{X} := \bar{\Omega}$ where Ω is an open bounded connected subset of $\mathbb{R}^d$ with $\mu(\partial\Omega) = 0$, that μ is equivalent to the Lebesgue measure on $\mathbf{X}$. Moreover for every $y \in \mathbf{Y}$ $c(\cdot, y)$ is differentiable with* *gradient* $\nabla_x c$ *bounded on* $\mathbf{X} \times \mathbf{Y}$, *and let* $\nu \in \mathscr{P}(\mathbf{Y})$, *then there exist a Kantorovich potential, unique up to addition of constants,* ϕ *between* μ *and* ν *and for every* $\rho \in \mathscr{P}(\mathbf{Y})$ *one has*

$$\lim_{\epsilon \to 0^+} \frac{\mathcal{W}_c(\mu, \nu - \epsilon(\nu - \rho)) - \mathcal{W}_c(\mu, \nu)}{\epsilon} = \int_Y \phi^c d(\rho - \nu).$$

Proof. We refer [13, Proposition 6.1] for the proof of existence and uniqueness of the Kantorovich potential.
From Theorem 1.1.9, there is one c-convex function ϕ such that

$$\mathcal{W}_c(\mu, \nu) = \int_{\mathbf{Y}} \phi^c(y) d\nu(y) - \int_{\mathbf{X}} \phi(x) d\mu(x).$$

Moreover $\phi = C$ is optimal for any constant C. Then, choosing one reference point $x_0 \in \mathbf{X}$, we choose C such that $\phi(x_0) = 0$. As ϕ is c-convex, but might not be optimal for $\mathcal{W}_c(\mu, \nu_\epsilon)$, where $\nu_\epsilon := \nu - \epsilon(\nu - \rho)$, we have

$$\begin{aligned} \mathcal{W}_c(\mu, \nu_\epsilon) &\geq \int_{\mathbf{Y}} \phi^c(y) d\nu_\epsilon(y) - \int_{\mathbf{X}} \phi(x) d\mu(x) \\ &= \int_{\mathbf{Y}} \phi^c(y) d\nu(y) + \epsilon \int_{\mathbf{Y}} \phi^c(y) d\rho - \epsilon \int_{\mathbf{Y}} \phi^c(y) d\nu - \int_{\mathbf{X}} \phi(x) d\mu(x). \end{aligned}$$

Hence,

$$\frac{\mathcal{W}_c(\mu, \nu_\epsilon) - \mathcal{W}_c(\mu, \nu)}{\epsilon} \geq \int_{\mathbf{Y}} \phi^c d(\rho - \nu).$$

Similarly, for every $\epsilon \in [0, 1]$ let ϕ_ϵ, with $\phi_\epsilon(x_0) = 0$, be the Kantorovich potential between μ and ν_ϵ. Then we have

$$\frac{\mathcal{W}_c(\mu, \nu_\epsilon) - \mathcal{W}_c(\mu, \nu)}{\epsilon} \leq \int_{\mathbf{Y}} \phi_\epsilon^c d(\rho - \nu).$$

As c is uniformly continuous and bounded, there is a modulus of continuity (a positive non decreasing function with $\omega(0) = 0$) $\omega : \mathbb{R} \to \mathbb{R}$, such that

$$|c(x, y) - c(x', y')| \leq \omega(\|x - x'\| + \|y - y'\|).$$

Then, by the definition of c-convexity, we have that

$$|\phi_\epsilon(x) - \phi_\epsilon(x')| \leq \omega(\|x - x'\|)$$

for each $\epsilon \in (0, 1)$. Hence $\{\phi_\epsilon\}_{\epsilon \in [0,1]}$ is bounded and equicontinuous, and applying the Arzela Ascoli Theorem, up to sub-sequence of functions it convergence uniformly to some continuous function $\bar{\phi}$. Then by Theorem 1.1.10 we know that $\nu \mapsto \mathcal{W}_c(\mu, \nu)$ is continuous with respect the weak convergence. Then for an arbitrary $\alpha > 0$, there exist $\epsilon^* > 0$ such that, for every $x \in \mathbf{X}$,

$$-\alpha + \bar{\phi}(x) < \phi_\epsilon(x) < \alpha + \bar{\phi}(x),$$

whenever $0 < \epsilon < \epsilon^*$. Given an arbitrary $y \in \mathbf{Y}$, one has

$$(\forall x \in \mathbf{X}) \qquad -\alpha + \bar{\phi}(x) + c(x, y) < \phi_\epsilon(x) + c(x, y) < \alpha + \bar{\phi}(x) + c(x, y).$$

Taking the infimum over all x in the latter inequality we conclude that

$$(\forall y \in \mathbf{Y}) \qquad -\alpha + \inf_{x\in\mathbf{X}}(\bar{\phi}(x) + c(x,y)) < \phi^c_\epsilon(y) < \alpha + \inf_{x\in\mathbf{X}}(\bar{\phi}(x) + c(x,y)).$$

In other words, up to a subsequence, $\{\phi^c_\epsilon\}_{\epsilon\in[0,1]}$ converges uniformly to $(\bar{\phi})^c$. Then by Dominated Convergence we have

$$\begin{aligned}\mathcal{W}_c(\mu,\nu) = \limsup_{\epsilon\to 0^+}\mathcal{W}_c(\mu,\nu_\epsilon) &= \limsup_{\epsilon\to 0^+}\left(\int_{\mathbf{Y}}\phi^c_\epsilon(y)d\nu_\epsilon(y) - \int_{\mathbf{X}}\phi_\epsilon(x)d\mu(x)\right)\\ &= \left(\int_{\mathbf{Y}}\bar{\phi}^c(y)d\nu(y) - \int_{\mathbf{X}}\bar{\phi}(x)d\mu(x)\right).\end{aligned}$$

Then it is observed that $\bar{\phi}$ is a Kantorovich Potential, with $\bar{\phi}(x_0) = 0$, so by the uniqueness of ϕ, ϕ_ϵ converges uniformly to ϕ as $\epsilon \to 0^+$. This implies that

$$\limsup_{\epsilon\to 0^+}\frac{\mathcal{W}_c(\mu,\nu_\epsilon) - \mathcal{W}_c(\mu,\nu)}{\epsilon} \leq \int_{\mathbf{Y}}\phi^c d(\rho-\nu).$$

□

Finally if we assume that $\mathcal{V}[\nu] := \frac{\delta\mathcal{E}}{\delta\nu}$, on $\mathcal{D}$, we can consider the variational problem

$$\inf_{\nu\in\mathcal{D}}\mathcal{J}_\mu[\nu], \tag{2.2.2}$$

where $\mathcal{J}_\mu[\nu] := \mathcal{W}_c(\mu,\nu) + \mathcal{E}[\nu]$.
We now state one of the main results of Carlier and Blanchet [8, Theorem 3.2].

Theorem 2.2.8. *Assume that $\mathcal{V}$ is the differential of $\mathcal{E}$ (in the sense of Definition 2.2.5), and all the assumptions in Lemma 2.2.7 hold true. If ν solves the variational equation* (2.2.2)*, and γ is an optimal transport plan between μ and ν, then γ is a Cournot-Nash equilibrium.*

Proof. Let ν be a minimizer of equation (2.2.2) and let us take $\rho \in \mathcal{D}$ and $\epsilon \in (0,1)$. Then we have

$$\frac{\mathcal{J}_\mu[\nu + \epsilon(\rho-\nu)] - \mathcal{J}_\mu[\nu]}{\epsilon} \geq 0.$$

Furthermore, by Lemma 2.2.7,

$$\int_{\mathbf{Y}}(\phi^c + \mathcal{V}[\nu])d\rho \geq \int_{\mathbf{Y}}(\phi^c + \mathcal{V}[\nu])d\nu.$$

Set $M := \operatorname{ess\,inf}_{y\in\mathbf{Y}}\{\phi^c(y) + \mathcal{V}[\nu](y)\}$ (with respect measure ν), then the next equality will hold ν a.e.

$$\phi^c + \mathcal{V}[\nu] = \inf_{\rho\in\mathcal{D}}\int_{\mathbf{Y}}(\phi^c + \mathcal{V}[\nu])d\rho = M.$$

Since γ is an optimal transport, we have that $\phi^c + \phi = c\ \gamma$ a.e. and $\phi^c + \phi \geq c$ pointwise. We thus have

$$\begin{cases} c(x,y) + \mathcal{V}(y)[\nu] \geq M + \phi(x) & \text{for all } x \in \mathbf{X} \text{ and } \mu_0\text{-a.e. y} \in \mathbf{Y};\\ c(x,y) + \mathcal{V}(y)[\nu] = M + \phi(x) & \text{for all } \gamma - (x,y).\end{cases}$$

□

2.3 From Nash to Cournot-Nash

This section deals with the notion of limits of classical Nash equilibria, when the number of payers tends to infinity. Most of the Section is based on the approach by Larsly and Lions [34] and its generalization by Carlier and Blanchet [7].

Let $\mathbf{X}$ and Θ be compact metric spaces with corresponding distances $d_{\mathbf{X}}$ and d_Θ, and let $\Theta_N := \{\theta_1, \dots, \theta_N\}$ be a finite subset of the type space Θ. As the number of agents is finite we only are interested in their types. Also note that some agents could belong to the same type. We are going to consider a N- person game where all the agents have the same strategy space X. We assume that the cost of agent i, depends on his type $\theta_i \in \Theta_N$, his strategy $x \in \mathbf{X}$, and his symmetry with respect to the others agents strategies x_{-i}, that is,

$$J_i^N(x_i, x_{-i}) = J^N(\theta_i, x_i, x_{-i}) = J^N(\theta_i, x_i, (x_{\sigma(j)})_{j\neq i}) \quad \text{for every} \quad \sigma \in S_i^{N-1},$$

where S_i^{N-1} denotes the set of permutations of $\{1, \dots, N\} \setminus \{i\}$. Moreover, we will to assume that there exists a modulus of continuity ω such that for every N, every $(\theta_i, \theta_j) \in \Theta_N \times \Theta_N$, every (x_i, x_{-i}) and (y_i, y_{-i}) in $\mathbf{X}^N$, we have

$$\begin{aligned}|J^N(\theta_i, x_i, x_{-i}) - J^N(\theta_j, y_i, y_{-i})| &\leq \omega(d_\Theta(\theta_i, \theta_j)) + \omega(d_{\mathbf{X}}(x_i, y_i)) \\ &+\omega\left(\mathcal{W}_1\left(\frac{1}{N-1}\sum_{j\neq i}\delta_{\theta_{x_j}}, \frac{1}{N-1}\sum_{j\neq i}\delta_{\theta_{y_j}}\right)\right),\end{aligned}$$

where $\mathcal{W}_1$ denotes the 1-Wasserstein metric. Note that

$$\mathcal{W}_1\left(\frac{1}{N-1}\sum_{j\neq i}\delta_{\theta_{x_j}}, \frac{1}{N-1}\sum_{j\neq i}\delta_{\theta_{y_j}}\right) = \min_{\sigma\in S_{N-1}} \frac{1}{N-1}\sum_{j\neq i} d_{\mathbf{X}}(x_j, y_{\sigma(j)}).$$

As we want to construct Cournot-Nash equilibria by taking a limit, we need to be congruent about the dimension where the objects of study are defined, so we need to extend J^N to $\Theta \times \mathbf{X} \times \mathscr{P}(\mathbf{X})$, through the cost F^N given by

$$F^N(\theta, x, \nu) := \inf_{(x_{-i},\theta_i)\in \mathbf{X}^{N-1}\times\Theta_N} \left\{ J^N(\theta_i, x, x_{-i}) + \omega(d_\Theta(\theta_i, \theta)) + \omega\left(\mathcal{W}_1\left(\nu, \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j}\right)\right)\right\}.$$

This is known as the classical McShane construction, and it extends the cost function to $\Theta \times \mathbf{X} \times \mathscr{P}(\mathbf{X})$ and it preserves the Nash equilibria for the N-game. Moreover, F^N is a sequence of uniformly equicontinuous functions on $\Theta \times \mathbf{X} \times \mathscr{P}(\mathbf{X})$.

Theorem 2.3.1 (Pure Nash equilibria converge to Cournot-Nash equilibria). *Let $\bar{x}^N = (x_1^N, \dots, x_N^N)$ be a Nash equilibrium for the game above, and define*

$$\mu^N := \frac{1}{N}\sum_{i=1}^N \delta_{\theta_i}, \quad \nu^N := \frac{1}{N}\sum_{i=1}^N \delta_{x_i^N}, \quad \textit{and } \gamma^N := \frac{1}{N}\sum_{i=1}^N \delta_{(\theta_i, x_i^N)}.$$

In addition, assume that, up to the extraction of subsequences,

$$\mu^N \overset{*}{\rightharpoonup} \mu, \quad \nu^N \overset{*}{\rightharpoonup} \nu, \quad \gamma^N \overset{*}{\rightharpoonup} \gamma \ \textit{and}\ F^N \to F \in C_b(\Theta \times \mathbf{X} \times \mathscr{P}(\mathbf{X})).$$

Then γ is a Cournot-Nash equilibrium for F and μ.

Proof. First, we set

$$\bar{\nu}_i^N := \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j^N}.$$

Now, we estimate the distance between ν^N and $\bar{\nu}_i^N$;

$$\mathcal{W}_1(\nu^N,\bar{\nu}_i^N) = \inf_{\lambda\in\Pi(\nu^N,\bar{\nu}_i^N)}\int_{\mathbf{X}\times\mathbf{X}} d_{\mathbf{X}}(x,y)d\lambda(x,y).$$

In particular, note that the product measure generated by ν^N and $\bar{\nu}_i^N$ is in $\Pi(\nu^N,\bar{\nu}_i^N)$ as $d_X(\cdot,\cdot)\geq 0$ we can use the Tonelli's theorem to obtain

$$\inf_{\lambda\in\Pi(\nu^N,\bar{\nu}_i^N)}\int_{\mathbf{X}\times\mathbf{X}} d_{\mathbf{X}}(x,y)d\lambda(x,y) \leq \int_{\mathbf{X}}\int_{\mathbf{X}} d_{\mathbf{X}}(x,y)d\nu^N(x)d\bar{\nu}_i^N(y) = \frac{1}{N(N-1)}\sum_{k=1}^{N}\sum_{j\neq i} d_{\mathbf{X}}(x_j,x_k).$$

There are at least $N-1$ terms in the sum that are 0 and for every x, y in $\mathbf{X}$, $d_{\mathbf{X}}(x,y)\leq \mathrm{Diam}\,(\mathbf{X})$ / Therefore

$$\mathcal{W}_1(\nu^N,\bar{\nu}_i^N) \leq \frac{\mathrm{Diam}\,(\mathbf{X})}{N}.$$

As $\bar{x}^N$ is a Nash equilibrium we have

$$(\forall y\in\mathbf{X})\quad F^N(\theta_i,x_i^N,\bar{\nu}_i^N)\leq F^N(\theta_i,y,\bar{\nu}_i^N),$$

Hence, we obtain

$$(\forall y\in\mathbf{X})\quad F^N(\theta_i,x_i^N,\nu^N)\leq F^N(\theta_i,y,\nu^N)+\epsilon_N,$$

where $\epsilon_N = 2\omega(\mathrm{Diam}\,(\mathbf{X})\,/N)$. Summing over i and dividing by N, we have

$$\int_{\Theta\times\mathbf{X}} F^N(\theta,x,\nu^N)d\gamma^N(\theta,x) \leq \int_{\Theta}\min_{y\in\mathbf{X}} F^N(\theta,y,\nu^N)d\mu^N(\theta)+\epsilon_N.$$

Because the equicontinuity of F^N and the existence of the limit in $C_b(\Theta,\mathbf{X},\mathscr{P}(\mathbf{X}))$, letting $N\to\infty$, we obtain

$$\int_{\Theta\times\mathbf{X}} F(\theta,x,\nu)d\gamma(\theta,x) \leq \int_{\Theta}\min_{y\in\mathbf{X}} F(\theta,y,\nu)d\mu(\theta).$$

As γ has marginals μ and ν, we have deduced that γ is a Cournot-Nash equilibrium, for F and μ. □

As it was mentioned in Section 3.1, the conditions that allow the existence of pure Nash equilibria are strong, and do not include some interesting cases of study. This is why we now consider the mixed strategy extension which allows, in this case, the existence of Nash equilibria.

Theorem 2.3.2. *If the assumption about modulus of continuity holds for $\omega(t) = Kt$, then the conclusion of Theorem 2.3.1 applies to the extension in mixed strategies with the extended cost*

$$\bar{J}^N(\theta_i, \pi_i, \pi_{-i}) := \int_{\mathbf{X}^N} J^N(\theta_i, x, x_{-i}) d\pi_i(x) \otimes_{j\neq i}^N d\pi_j(x_j).$$

Proof. Take θ_i and θ_j in Θ. By hypothesis,

$$|J^N(\theta_i, x_i, x_{-i}) - J^N(\theta_j, x_i, x_{-i})| \leq K d_\Theta(\theta_i, \theta_j).$$

Integration of above equation yields

$$|\bar{J}^N(\theta_i, \pi_i, \pi_{-i}) - \bar{J}^N(\theta_j, \pi_i, \pi_{-i})| \leq K d_\Theta(\theta_i, \theta_j).$$

Let us take η_i, $\pi_i \in \mathscr{P}(\mathbf{X})$, and let $\Gamma \in \Pi(\eta_i, \pi_i)$ be such that

$$\mathcal{W}_1(\pi_i, \eta_i) = \int_{\mathbf{X}\times\mathbf{X}} d_{\mathbf{X}}(x, y) d\Gamma(x, y).$$

Then, by definition, we have

$$\begin{aligned}
&\left|\bar{J}^N(\theta_i, \pi_i, \pi_{-i}) - \bar{J}^N(\theta_i, \eta_i, \pi_{-i})\right| \\
&= \left|\int_{\mathbf{X}^{N-1}} \left(\int_{\mathbf{X}} J^N(\theta_i, x, \pi_{-i}) d\pi_i(x) - \int_{\mathbf{X}} J^N(\theta_i, y, \pi_{-i}) d\eta_i(y)\right) \otimes_{j\neq i} d\pi_j(x_j)\right| \\
&= \left|\int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}\times\mathbf{X}} \left(J^N(\theta_i, x, \pi_{-i}) d\pi_i(x) - J^N(\theta_i, y, \pi_{-i})\right) d\pi_i(x) d\eta_i(y) \otimes_{j\neq i} d\pi_j(x_j)\right| \\
&\leq \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}\times\mathbf{X}} \left|J^N(\theta_i, x, \pi_{-i}) d\pi_i(x) - J^N(\theta_i, y, \pi_{-i})\right| d\pi_i(x) d\eta_i(y) \otimes_{j\neq i} d\pi_j(x_j).
\end{aligned}$$

As Γ has marginals π_i and η_i, it follows that

$$\begin{aligned}
&\left|\bar{J}^N(\theta_i, \pi_i, \pi_{-i}) - \bar{J}^N(\theta_i, \eta_i, \pi_{-i})\right| \\
&\leq \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}\times\mathbf{X}} \left|J^N(\theta_i, x, \pi_{-i}) d\pi_i(x) - J^N(\theta_i, y, \pi_{-i})\right| d\Gamma(x, y) \otimes_{j\neq i} d\pi_j(x_j) \\
&\leq \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}\times\mathbf{X}} K d_{\mathbf{X}}(x, y) d\Gamma(x, y) \otimes_{j\neq i} d\pi_j(x_j) \\
&= \int_{\mathbf{X}^{N-1}} \mathcal{W}_1(\pi_i, \eta_i) \otimes_{j\neq i} d\pi_j(x_j) = K\mathcal{W}_1(\pi_i, \eta_i).
\end{aligned}$$

Let us take $\sigma \in S_i^{N-1}$ and $\Gamma_j \in \Pi(\pi_j, \eta_{\sigma(j)})$ for $j \neq i$ in such a way that

$$\mathcal{W}_1(\pi_j, \eta_{\sigma(j)}) = \int_{\mathbf{X}\times\mathbf{X}} d_{\mathbf{X}}(x, y) d\Gamma_j(x, y).$$

Then, by symmetry,

$$\begin{aligned}
&\left|\bar{J}^N(\theta_i, \pi_i, \pi_{-i}) - \bar{J}^N(\theta_i, \pi_i, \eta_{-i})\right| = \left|\bar{J}^N(\theta_i, \pi_i, \pi_{-i}) - \bar{J}^N(\theta_i, \pi_i, \{\eta_{\sigma(j)}\}_{j\neq i})\right| \\
&= \left|\int_{\mathbf{X}^N} J^N(\theta_i, x_i, x_{-i}) d\pi_i(x_i) \otimes_{j\neq i} d\pi_j(x_j) - \int_{\mathbf{X}^N} J^N(\theta_i, x_i, y_{-i}) d\pi_i(x_i) \otimes_{j\neq i} d\eta_{\sigma(j)}(y_j)\right| \\
&= \left|\int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}^N} J^N(\theta_i, x_i, x_{-i}) d\pi_i(x_i) \otimes_{j\neq i} d\pi_j(x_j) \otimes_{j\neq i} d\eta_{\sigma(j)}(y_j)\right. \\
&\left. - \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}^N} J^N(\theta_i, x_i, y_{-i}) d\pi_i(x_i) \otimes_{j\neq i} d\eta_{\sigma(j)}(y_j) \otimes_{j\neq i} d\pi_j(x_j)\right|.
\end{aligned}$$

By Fubini's theorem, we have

$$
\begin{aligned}
&= \left| \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}^N} J^N(\theta_i, x_i, x_{-i}) - J^N(\theta_i, x_i, y_{-i}) d\pi_i(x_i) \otimes_{j \neq i} d\pi_j(x_j) \otimes_{j \neq i} d\eta_{\sigma(j)}(y_j) \right| \\
&\leq \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}^N} \left| J^N(\theta_i, x_i, x_{-i}) - J^N(\theta_i, x_i, y_{-i}) \right| d\pi_i(x_i) \otimes_{j \neq i} d\pi_j(x_j) \otimes_{j \neq i} d\eta_{\sigma(j)}(y_j), \\
&\leq \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}^N} \left| J^N(\theta_i, x_i, x_{-i}) - J^N(\theta_i, x_i, y_{-i}) \right| d\pi_i(x_i) \otimes_{j \neq i} \{ d\pi_j(x_j) \times d\eta_{\sigma(j)}(y_j) \},
\end{aligned}
$$

since, for $j \neq i$, Γ_j has marginals π_j and η_j, respectively,

$$
\begin{aligned}
&= \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}^N} \left| J^N(\theta_i, x_i, x_{-i}) - J^N(\theta_i, x_i, y_{-i}) \right| d\pi_i(x_i) \otimes_{j \neq i} d\Gamma_j(x_j, y_j) \\
&\leq \int_{\mathbf{X}^{N-1}} \int_{\mathbf{X}^N} \frac{K}{N-1} \sum_{j \neq i} d_{\mathbf{X}}(x_j, y_{\sigma(j)}) d\pi_i(x_i) \otimes_{j \neq i} d\Gamma_j(x_j, y_j) \\
&\leq \frac{K}{N-1} \sum_{j \neq i, j=1}^{N} \mathcal{W}_1(\pi_j, \eta_{\sigma(j)}).
\end{aligned}
$$

Because σ is arbitrary, we have

$$
\left| \bar{J}^N(\theta_i, \pi_i, \pi_{-i}) - \bar{J}^N(\theta_i, \pi_i, \eta_{-i}) \right| \leq K \min_{\sigma \in S_i^{N-1}} \frac{1}{N-1} \sum_{j \neq i, j=1}^{N} \mathcal{W}_1(\pi_j, \eta_{\sigma(j)}).
$$

We conclude with the observation that

$$
\begin{aligned}
\left| \bar{J}^N(\theta_i, \pi_i, \pi_{-i}) - \bar{J}^N(\theta_j, \eta_i, \eta_{-i}) \right| &\leq K d_\Theta(\theta_i, \theta_j) + K \mathcal{W}_1(\pi_i, \eta_i) \\
&\quad + K \min_{\sigma \in S_i^{N-1}} \frac{1}{N-1} \sum_{j \neq i, j=1}^{N} \mathcal{W}_1(\pi_j, \eta_{\sigma(j)}).
\end{aligned}
$$

That is the extension of J still satisfies the equi-continuity assumption, with the same modulus of continuity, so we can apply Theorem 2.3.1. □

Concluding remarks

In this chapter we have founded suitable assumption under a Cournot-Nash equilibrium is given in terms of an optimal transference plan. This variational approach is due to A. Blanchet and G. Carlier. Also we remark that this notion of limit of Nash equilibria is reminiscent of the work of P. Lions [34]. In this thesis we work with a more general model, when we take into consideration the type variable, this approach is due to G. Carlier and A. Blanchet [7].

As we have mentioned in the previous chapter, in applications, is relevant to know whether or not there exists a Nash equilibrium for a given game G. Moreover, is of high relevancy to explicitly calculate that equilibrium for such game. Nevertheless, the complexity of calculating those equilibria grows exponentially. In particular, 2.3.1 and 2.3.2 ensemble with 2.2.8 give us a bidirectional way for calculating those equilibria, i.e., given a game G with $N \in \mathbb{N}$ players, with a sufficiently large N, then a Nash equilibrium for the game G can be approximated by a Cournot-Nash equilibria for the limit game, vice-versa, if we can find a suitable sequence of $N-$person games, with $N \in \mathbb{N}$, then the Cournot-Nash equilibria is given by the limit of Nash equilibria for the $N-$person game.

Chapter 3

Optimal Transport and Optimal Control

In this chapter, we address a different –but not to apart from the traditional – approach for solving *optimal control problems*, which mainly consist of a set of admissible actions U, a set of states $\mathbf{X}$, and a non linear dynamic T–which describes the evolution of the state through time and the actions taken–. We concern to minimise the total cost associated to a sequence of actions taken. We aim to give this optimal sequence of actions in terms of optimal tranference plans. For the big picture about discrete-time optimal control see [32, 30, 29].

Section 1 deals with the analysis of a dynamical system, i.e., we analyse if a dynamical system could be control through the time to a given fixed structure, in literature this is known as the controllability analysis. Section 2 gives an approach for solvability to the discrete-time optimal control problem; given a suitable dynamical system, we find an optimal path of actions that the controller must follow in order to optimize his objective function. This optimal path is given in term of a sequence of optimal transport plans.

3.1 Controllability

Let $\mathbf{X}$ be a finite-dimensional manifold. For the purpose of this thesis it suffices to consider, for example,$\mathbf{X} = \mathbb{R}^d$, Let U denote the set of actions. We will assume that U is a compact subset of a metric space. A measurable function $V := \mathbf{X} \to U$ is called *feedback* –or *deterministic Markov*– control law.

Definition 3.1.1. *a) For a set $M \subset \mathbf{X}$ and $p \in \mathbb{N}$, we define the set*

$$D_M^p := \left\{ \sum_{k=1}^{p} c_k \delta_{y_k};\ \ \{y_k\}_{k=1}^p \subset M,\ \ \{c_k\}_{k=1}^p \subset [0,1],\ \ \sum_{k=1}^{p} c_k = 1 \right\}. \tag{3.1.1}$$

We will also define $D_M = \cup_{p=1}^{\infty} D_M^p$.

b) We define $\mathcal{Y}(\mathbf{X}, U)$ the sets of maps $K : \mathbf{X} \times \mathcal{B}(U) \to \mathbb{R}$, where $K(\cdot, A)$ is a $\mathcal{B}(X)$ measurable function for each $A \in \mathcal{B}(U)$ and $K(x, \cdot) \in \mathscr{P}(U)$ for each

$x \in \mathbf{X}$. A map $K \in \mathcal{Y}(\mathbf{X}, U)$ is called a stochastic–or randomised– feedback law.

Now, let us consider the nonlinear discrete time control system

$$\begin{cases} x_{n+1} = T(x_n, u_n) & (\forall n \in \mathbb{N}) \\ x_0 \in \mathbf{X}, \end{cases} \tag{3.1.2}$$

where $\{x_n\}_{n=0}^{\infty} \subset \mathbf{X}$, $\{u_n\}_{n=0}^{\infty} \subset U$, and $T : \mathbf{X} \times U \to \mathbf{X}$ is a continuous map. Then there is an induced control system over the probability measures in $\mathscr{P}(\mathbf{X})$, given by

$$\begin{cases} \mu_{n+1} = T(\cdot, u_n)_{\#}\mu_n & (\forall n \in \mathbb{N}), \\ \mu_0 \in \mathscr{P}(\mathbf{X}). \end{cases} \tag{3.1.3}$$

Inspired by the Monge problem, our aim here is to find some deterministic path, i.e., to find a sequence of feedback laws $v_n : \mathbf{X} \to U$, $n \in \{0, 1, \ldots, N-1\}$, such that, given some $N \in \mathbb{N}$ and a target measure $\mu^f \in \mathscr{P}(\mathbf{X})$, the system satisfies

$$\begin{cases} \mu_{n+1} = T(\cdot, v_n(x))_{\#}\mu_n & n \in \{0, 1, \ldots, N-1\}, \\ \mu_N = \mu^f. \end{cases}$$

Note that the formulation above stated presents the same limitations as in Monge problem; we can not split one Dirac delta into two Dirac deltas with a measurable map. Hence, inspire by the relaxation of Monge Problem, introduced in Chapter 1 by the name of Monge-Kantorovich problem, our aim is to find some stochastic feedback laws instead of searching deterministic ones. So, given some $N \in \mathbb{N}$ and a target measure $\mu^f \in \mathscr{P}(\mathbf{X})$, determine if there are ,or not, a sequence of stochastic feedback laws $K_n \in \mathcal{Y}(\mathbf{X}, U)$, $n \in \{0, 1, \ldots, N-1\}$, such that the system satisfies

$$\begin{cases} \mu_{n+1} = T_{\#}^{cl,n}\mu_n & n \in \{0, 1, \ldots, N-1\}, \\ \mu_N = \mu^f, \end{cases} \tag{3.1.4}$$

where, for each $A \in \mathcal{B}(\mathbf{X})$ and $n \in \{0, 1, \ldots, N-1\}$,

$$T_{\#}^{cl,n}\mu_n(A) := \int_{\mathbf{X}} \int_U \mathbf{1}_A(T(x, u)) K_n(x, du) d\mu_n(x). \tag{3.1.5}$$

Definition 3.1.2 (Reachable states)**.** *For some $x \in \mathbf{X}$, let $R_1^x := \{T(x, u); u \in U\}$ be the set of reachable states in one step from x. Then inductively we define the set*

$$R_m^x = \cup_{y \in R_{m-1}^x} \{T(y, u); u \in U\},$$

of reachable states from x in m steps, with $m \geq 2$.

Now instead of proving that there is always a $\{K_n\}_{n=0}^{N-1} \subset \mathcal{Y}(\mathbf{X}, U)$ that satisfies (3.1.4), we will consider to look for $N-1$ measures $\nu_n \in \mathscr{P}(\mathbf{X} \times U)$, $n \in \{0, 1, \ldots, N-1\}$, such that, given an initial measure $\mu_0 \in \mathscr{P}(\mathbf{X})$ and a target measure $\mu^f \in \mathscr{P}(\mathbf{X})$, the measures $\{\nu_n\}_{n=0}^{N-1}$ satisfy

$$\mu_{n+1} = T_{\#}\nu_n, \qquad n \in \{0, 1, \ldots, N-1\}, \tag{3.1.6}$$

with $\nu_n(A \times U) = \mu_n(A)$ for all $A \in \mathcal{B}(\mathbf{X})$ and $\mu_N = \mu^f$.

Proposition 3.1.3. *Let $\mu_0 = \delta_{x_0}$ for some $x_0 \in \mathbf{X}$. Let $\mu^f \in D^p_M$ for a compact subset $M \subset \mathbf{X}$, some $p \in \mathbb{N}$, such that the Spt$\{\mu^f\} \subset R^{x_0}_N$. Then there exist a sequence of measures, $\{\nu_n\}_{n=0}^{N-1} \in \mathscr{P}(\mathbf{X} \times U)$ such that*

$$\mu_{n+1} = T_\# \nu_n, \qquad n \in \{0, 1, \ldots, N-1\},$$

with $\nu_n(A \times U) = \mu_n(A)$ for all $A \in \mathcal{B}(\mathbf{X})$ and $\mu_N = \mu^f$.

Proof. Let $\mu^f := \sum_{i=1}^p c_i \delta_{y_i}$ as in (3.1.1), and note that Spt$\{\mu^f\} := \{y_1, \ldots y_p\}$, so by hypothesis, for each $i \in \{1, 2, \ldots, p\}$, there exist $\{u^i_n\}_{n=0}^{N-1} \subset U$, that satisfy (3.1.2), with $x^i_0 = x_0$ and $x^i_N = y_i$. Let us define, for each $i \in \{1, 2, \ldots, p\}$, $\nu^i_n := \delta_{(x^i_{n-1}, u^i_n)}$. Note that $T_\# \nu^i_n = \delta_{x^i_n}$. Then setting $\nu_n := \sum_{i=1}^p c_i \nu^i_n$, for all $n \in \{0, 1, \ldots, N-1\}$. Hence, we have that $T_\# \nu_n = \sum_{i=1}^p c_i T_\# \nu^i_n = \sum_{i=1}^p c_i \delta_{x^i_n} = \mu_{n+1}$ for all $n \in \{0, 1, \ldots, N-1\}$ and $T_\# \nu_{N-1} = \sum_{i=1}^p c_i T_\# \nu^i_{N-1} = \sum_{i=1}^p c_i \delta_{y^i} = \mu^f$. Finally

$$\nu_n(A \times U) = \sum_{i=1}^p c_i \delta_{(x^i_{n-1}, u^i_n)}(A \times U) = \sum_{i=1}^p c_i \delta_{x^i_{n-1}}(A) \delta_{u^i_n}(U) = \mu_n,$$

for $n \in \{0, 1, \ldots, N-1\}$ and every $A \in \mathcal{B}(\mathbf{X})$ □

Lemma 3.1.4. *Let $\mu_0 \in D^p_A$, and $\mu^f \in D^q_A$, with A a compact subset of $\mathbf{X}$ and some $p, q \in \mathbb{Z}_+$, such that Spt$\{\mu^f\} \subset R^x_N$ for each $x \in$ Spt$\{\mu_0\}$. Then there exists a sequence of measures, $\{\nu_n\}_{n=0}^{N-1} \in \mathscr{P}(\mathbf{X} \times U)$ such that*

$$\mu_{n+1} = T_\# \nu_n, \qquad n \in \{0, 1, \ldots, N-1\},$$

with $\nu_n(A \times U) = \mu_n(A)$ for all $A \in \mathcal{B}(\mathbf{X})$ and $\mu_N = \mu^f$.

Proof. Let $\mu_0 := \sum_{i=1}^p c_i \delta_{y_i}$ as in (3.1.1), then, by proposition 3.1.3, there exist for each $i \in \{1, 2, \ldots, p\}$ measures $\nu^i_n \in \mathscr{P}(\mathbf{X} \times U)$ such that if we set $\eta^i_0 := \delta_{y^i}$ they satisfy (3.1.2), with $\nu^i_n(A \times U) = \eta^i_n(A)$ for $n \in \{0, 1, \ldots, N-1\}$ and every $A \in \mathcal{B}(\mathbf{X})$, and $\eta^i_N = \mu^f$. So if we set $\nu_n := \sum_{i=1}^p c_i \nu^i_n$ for $n \in \{0, 1, \ldots, N-1\}$ the result follows. □

Proposition 3.1.5. *Let $\mu_0, \mu_f \in \mathscr{P}(\mathbf{X})$ with compact supports, such that the Spt$\{\mu_f\} \subset R^x_N$ for each $x \in$ Spt$\{\mu_0\}$. Then there exists a sequence of measures $\{\nu_n\}_{n=0}^{N-1} \in \mathscr{P}(\mathbf{X} \times U)$ such that*

$$\mu_{n+1} = T_\# \nu_n, \qquad n \in \{0, 1, \ldots, N-1\},$$

with $\nu_n(A \times U) = \mu_n(A)$ for all $A \in \mathcal{B}(\mathbf{X})$ and $\mu_N = \mu^f$.

Proof. Let $A_m := \cup_{x \in \text{Spt}\{\mu_0\}} R^x_m$. For $N = 1$ note that $A_1 = \{T(x, u); x \in \text{Spt}\{\mu_0\}, u \in U\}$. As the image of compact sets of continuous functions is compact, A_1 is compact. In general, if we suppose that A_n is compact, then $A_{n+1} := \{T(x, u); x \in A_n, u \in U\}$ is compact. In particular A_N is compact. From Theorem $A.0.11$ there exist $\{\mu^n_0\}_{n=1}^\infty$ and $\{\mu^n_f\}_{n=1}^\infty$ subsets of D_{A_N} such that $\mu^n_0 \rightharpoonup \mu_0$ and $\mu^n_f \rightharpoonup \mu_f$. Then, by Lemma 3.1.4, there exist $\{\nu^n_k\}_{k=0}^{N-1} \subset \mathscr{P}(\mathbf{X})$ such that

$$\mu^n_{k+1} = T_\# \nu^n_k, \qquad k \in \{0, 1, \ldots, N-1\},$$

with $\nu_k^n(A \times U) = \mu_k(A)$ for all $A \in \mathcal{B}(\mathbf{X})$ and $\mu_N^n = \mu_f^n$, for every $n \in \mathbb{Z}_+$. Since the map T is continuous, by Theorem A.0.8, there is a limit measure ν_k such that $T_\# \nu_k^n \rightharpoonup \nu_k$, for $k \in \{0, 1, \dots, N-1\}$. Then the result follow by the continuity of T. □

Theorem 3.1.6. *Let $\mu_0, \mu_f \in \mathscr{P}(\mathbf{X})$ with compact supports, and such that the $Spt\{\mu_f\} \subset R_N^x$ for each $x \in Spt\{\mu_0\}$. Then there exist a sequence of stochastic feedback laws $\{K_n\}_{n=0}^{N-1} \subset \mathcal{Y}(\mathbf{X}, U)$ such that (3.1.4) is satisfied. Hence μ_f is a reachable measure from μ_0.*

Proof. Note that $\mathbf{X}$ and U are separable spaces, so $\mathcal{B}(\mathbf{X}) \otimes \mathcal{B}(U) = \mathcal{B}(\mathbf{X} \times U)$. Then given $\nu \in \mathscr{P}(\mathbf{X} \times U)$, by the disintegration theorem, there exist a measure $\mu \in \mathscr{P}(\mathbf{X})$ and a stochastic feedback law $K \in \mathcal{Y}(\mathbf{X}, U)$ such that

$$\int_{A \times B} d\nu(x, u) = \int_A \int_B K(x, du) d\mu(x)$$

for all $A \in \mathcal{B}(\mathbf{X})$ and $B \in \mathcal{B}(U)$. Then applying Proposition 3.1.5 to get $\{\nu_n\}_{n=0}^{N-1}$ as in (3.1.6), we can apply the disintegration theorem to each ν_n to obtain K_n with their corresponding μ_n for each $n \in \{0, 1, \dots, N-1\}$. □

3.2 Optimal Control

Suppose that $c : \mathbf{X} \times U \to \mathbb{R}$ is a continuous function. Given a final time $N \in \mathbb{Z}_+$, an initial measure $\mu_0 \in \mathscr{P}(\mathbf{X})$ and a target measure $\mu_f \in \mathscr{P}(\mathbf{X})$, our aim is to find a solution for the optimization problem

$$\min_{\left\{\{\mu_{m+1}\}_{m=0}^{N-1} \subset \mathscr{P}(\mathbf{X}), \{K_n\}_{n=0}^{N-1} \subset \mathcal{Y}(\mathbf{X}, U)\right\}} \sum_{n=0}^{N-1} \int_X \int_U c(x, u) K_n(x, du) d\mu_n(x) \tag{3.2.1}$$

subject to the constrains

$$\begin{cases} \mu_{n+1} = T_\#^{cl,n} \mu_n & n \in \{0, 1, \dots, N-1\}, \\ \mu_N = \mu_f. \end{cases} \tag{3.2.2}$$

This can be interpreted as to minimise over all admissible paths of measures, that have initial and final measures μ_0 and μ_f, respectively. The existence of an admissible path will depend, as in previous section, on the fact that μ_f is reachable from μ_0 in N steps.

Note that instead of searching those $K_n \in \mathcal{Y}(\mathbf{X}, U)$, $n \in \{0, 1, \dots N\}$, one can convexify the problem (3.2.1) and try to solve in the space $\mathscr{P}(\mathbf{X} \times U)$;

$$\min_{\left\{\{\mu_m\}_{m=0}^{N-1} \subset \mathscr{P}(\mathbf{X}), \{\nu_n\}_{n=0}^{N-1} \subset \mathscr{P}(\mathbf{X}, U)\right\}} \sum_{n=0}^{N-1} \int_{\mathbf{X} \times U} c(x, u) d\nu_n(x, u) \tag{3.2.3}$$

subject to the constrains

$$\begin{cases} \mu_{n+1} = T^{cl,n}_{\#}\nu_n & n \in \{0,1,\dots,N-1\}, \\ \Pi_{\mathbf{X}\#}\nu_n = \mu_n \\ \mu_N = \mu_f. \end{cases} \tag{3.2.4}$$

Now we are ready to state the main theorem of the chapter:

Theorem 3.2.1. *Let $\mu_0, \mu_f \in \mathscr{P}(\mathbf{X})$ with compact support, and such that the Spt$\{\mu_f\} \subset R^x_N$ for each $x \in$ Spt$\{\mu_0\}$. Then the optimisation problem* (3.2.3),(3.2.4) *has a solution $\{(\mu_{n+1}, \nu_n)\}_{n=0}^{N-1}$.*

Proof. From Theorem 3.1.6 we know that the set of measures that satisfy (3.2.4) is not empty. Then, as shown in the proof of Proposition 3.1.5, $A_m := \cup_{x\in \mathrm{Spt}\{\mu_0\}} R^x_m$ is compact, so no matter what the choice of ν_n, the support of the sequence of measures is contained in a compact set. Moreover Spt $\{\nu_n\} \subset A_n \times U$ for every $n \in \{0, 1, \dots N-1\}$. Since c is continuous, is bounded over $A_n \times U$. In particular, $\sum_{k=0}^{N-1} \int_{\mathbf{X}\times U} c(x,u) d\nu_k(x,u)$ is bounded from below in the set of admissible sequences of measures. Thus, by definition of infimum, there is a sequence of $\{(\nu^k_n, \mu^k_{n+1})_{n=0}^{N-1}\}_{k=1}^{\infty}$ such that

$$\inf_{\left\{\{\mu_m\}_{m=0}^{N-1}\subset\mathscr{P}(\mathbf{X}),\{\nu_n\}_{n=0}^{N-1}\subset\mathscr{P}(\mathbf{X},U)\right\}} \sum_{n=0}^{N-1} \int_{\mathbf{X}\times U} c(x,u)d\nu_n(x,u) = \lim_{k\to\infty} \sum_{n=0}^{N-1} \int_{\mathbf{X}\times U} c(x,u)d\nu^k_n(x,u),$$

and for each $k \in \mathbb{Z}_+$, $(\nu^k_n, \mu^k_{n+1})_{n=0}^{N-1}$ satisfy the restriction (3.2.4). Set the compact set $A := \cup_{n=0}^{N} R^x_n$. Then for every $n \in \{0, 1 \dots, N-1\}$ and every $\epsilon > 0$, $\nu_n(A \times U) = 1 > 1-\epsilon$ and $\mu_n(A) = 1 > 1-\epsilon$. Thus $\{(\nu^k_n, \mu^k_{n+1})_{n=0}^{N-1}\}_{k=1}^{\infty}$ is tight. Therefore, there exists a subsequence $\{(\nu^{k_j}_n, \mu^{k_j}_{n+1})_{n=0}^{N-1}\}_{j=1}^{\infty}$ that weakly converges to some sequence $\{\nu^*_n, \mu^*_{n+1}\}_{n=0}^{N-1}$. Then by definition of weak convergence

$$\inf_{\left\{\{\mu_m\}_{m=0}^{N-1}\subset\mathscr{P}(\mathbf{X}),\{\nu_n\}_{n=0}^{N-1}\subset\mathscr{P}(\mathbf{X},U)\right\}} \sum_{n=0}^{N-1} \int_{\mathbf{X}\times U} c(x,u)d\nu_n(x,u) = \sum_{n=0}^{N-1} \int_{\mathbf{X}\times U} c(x,u)d\nu^*_n(x,u).$$

Finally, by continuity of the map T, $\{\nu^*_n, \mu^*_{n+1}\}_{n=0}^{N-1}$ still satisfy the restriction (3.2.4). □

Concluding remarks

Although the chapter is called optimal transport and optimal control, we have not used any reference from Chapter 1. This is because, we have proved Theorem 3.2.1 by the same techniques of the Theorem 1.0.8. In fact, we can consider ν_n as a collection of optimal transference plan, but instead of taking the condition over the second marginal we ask for the condition that the map $T^{cl,n}$ push the measure ν_n to μ_{n+1}. Since $T^{cl,n}$ is not linear, Theorem 1.0.8 can not be applied inductively to find a solution the optimisation problem (3.2.3), (3.2.4).

Chapter 4

Optimal Transport and Population Games

In this chapter we retake the framework of displacement interpolation stated in Chapter 1. We aim to analysis systems with infinitely many agents from a dynamical approach. We suppose a system where the agents are characterised only by the strategies they are taking. Actions taken by a finite sets of agents are negligible for the system. Therefore, a state at the time t is given by a probability measures over the set of strategies. Intuitively, the measure of a set A represents the proportion of the population (the set of agents). that are taking strategies in A. Here, each player is assumed to make decision according to a stochastic process instead of making one shot decision–as in Chapter 2–. In this case, we consider a discrete strategy set $S = \{0, 1, \dots n\}$.

Note that when we consider the discrete topology, the space of probabilities $\mathscr{P}(S)$ is a simplex:

$$\mathscr{P}(S) = \left\{ (p_k)_{k=0}^n \in \mathbb{R}^n; \ \sum_{k=0}^{n} p_k = 1, \ p_k \geq 0, \ k \in S \right\}. \tag{4.0.1}$$

Definition 4.0.1. *The interior of $\mathscr{P}(S)$ will be denoted by $\mathscr{P}(S)_o$, so*

$$\mathscr{P}_o(S) := \left\{ (p_k)_{k=0}^n \in \mathbb{R}^n; \ \sum_{k=0}^{n} p_k = 1, \ p_k > 0, \ k \in S \right\}.$$

The players are influenced at the hour of taking strategies by the utility perceived by the population– when we consider the average utility, the dynamic is known as the *replicator dynamic*–, thus, the payoff function is independent of the identity of the player. These population games are called autonomous. For every $k \in S$, $F_k : \mathscr{P}(S) \to \mathbb{R}$ will represent the gain of taking strategy k, given the state of the game $p \in \mathscr{P}(S)$. For every $k \in S$, F_k is assumed to be continuous with respect the weak topology in $\mathscr{P}(S)$ and the usual topology in $\mathbb{R}$.

Definition 4.0.2 (Population Games Nash Equilibria)**.** *A population state $p^* \in \mathscr{P}(S)$ is a Nash equilibrium for the population game if for given $k \in S$*

$$p_k^* > 0 \qquad \Rightarrow \qquad (\forall j \in S) \quad F_k(p^*) \geq F_j(p^*). \tag{4.0.2}$$

We also define de noisy payoff function $\bar{F}_k(p) := F_k(p) - \beta \log p_k$, $k \in S$.

Intuitively, the noisy payoff function models the risk-taking of the players; the fewer players currently select strategy k, the more likely a player is willing to take risk by switching to strategy k.

This chapter mainly focus on the Fokker-Planck equation. For each $x \in \mathbb{R}^d$, it has the form

$$\frac{\partial}{\partial t}\rho + \dot{\bigtriangledown}(f(x)\rho) = \beta \dot{\bigtriangledown} (AA^T \bigtriangledown \rho),$$

where $AA^T = A(x)A(x)^T$ is a nonnegative definite (diffusion) matrix and $f(x) \in \mathbb{R}^d$ is a (drift) vector function on x. Here, for each $t \geq 0$, the unknown is the probability density $\rho(t, x)$. Indeed, the Fokker-Planck equation describes the evolution of the transition probability of the solution to the stochastic differential equation

$$dX_t = f(X_t)dt = \sqrt{2\beta}A(X_t)dW_t,$$

where W_t is a $d-$dimensional standard Wiener process (so-called Brownian motion)

$$\int_A \rho(t, x)dx; = \int_A \mathbb{P}(X_t \in dx|X_0),$$

for each $x \in \mathbb{R}^d$, $t > 0$ and $A \in \mathcal{B}(\mathbb{R}^d)$(the Borel $\sigma-$algebra of $\mathbb{R}^d$).

4.1 Gradient Flows

In this section, in analogy to the 2-Wasserstein distance $\mathcal{W}_2(\mu, \nu)$, for μ, ν in $\mathscr{P}(\mathbb{R}^d)$, we aim to construct a metric over $\mathscr{P}(S)$ that will help us to describe the dynamics of the population. This time-dependent metric will help us to relate the optimal transport with the Fokker-Planck equation. We have mentioned in previous chapters that the notion of potential games is important to approach new optimization problems, and here is not the exception;

Definition 4.1.1. *A population game is named a potential game, if there exist a differentiable function* $\mathcal{F} : \mathscr{P}(S) \to \mathbb{R}$, *such that* $\frac{\partial}{\partial p_k}\mathcal{F}(p) = F_k(p)$ *for all* $k \in S$ *and* $p \in \mathscr{P}(S)$. *In this case* $\mathcal{F}$ *is called a potential function.*

Usually, every player in the game is allowed to take any strategy from S, but in some cases of interest, if one player has taken one strategy $j \in S$, he/she can not change arbitrarily to other strategy. This information is encoded in one graph, the strategy graph $G := (S, E)$ where S denotes the set of vertices, and $E \subset S \times S$ is the set of edges of the graph, where $\{i, j\} \in E$ means that a player is allowed to switch from strategy $i \in S$ to strategy $j \in S$. Note that the fact that one edge e belongs to E does not depend on the order of the strategies, i.e., the edges are bidirectional; thus we are not dealing with directed edges; fenote the neighborhood of $j \in S$ by $N(j) := \{k \in S; \quad \{j, k\} \in E\}$.

For any given strategy graph G, we can introduce an optimal transport metric on the simplex $\mathscr{P}(S)$.

Definition 4.1.2. *Given a function $\Phi : S \to \mathbb{R}$, define $\nabla\Phi : S \times S \to \mathbb{R}$ as*

$$\nabla\Phi_{j,k} = \begin{cases} \Phi_j - \Phi_k & \text{if } \{j,k\} \in E; \\ 0 & \text{otherwise.} \end{cases} \tag{4.1.1}$$

We say that a function $m := S \times S \to \mathbb{R}$ is anti-symmetric if $m_{i,j} = -m_{j,i}$. The divergence of m, denoted as $div(m) : S \to \mathbb{R}$, is defined by

$$div(m)_j = -\sum_{k \in N(j)} m_{j,k}, \qquad j \in S. \tag{4.1.2}$$

Given a strategy graph $G := (S, E)$, a set of continuous pay off functions $F_k : \mathscr{P}S \to \mathbb{R}$, for $k \in S$, and a population state $p \in \mathscr{P}(S)$. Then we can assign a weight to each edge $e \in E$ with a symmetric function $g : E \to \mathbb{R}$ defined by

$$g_{j,k}(p) := \begin{cases} p_k & \bar{F}_k(p) < \bar{F}_j(p); \\ p_j & \bar{F}_k(p) > \bar{F}_j(p); \\ \frac{p_k + p_j}{2} & \bar{F}_k(p) = \bar{F}_j(p). \end{cases}$$

This functions models the intensity with which an agent changes between strategies j and k. Indeed, this can be thought as the $Q-$matrix of a markovian processes.

We can now define the discrete inner product of $\nabla\Phi$ on $\mathscr{P}_o(S)$

$$\langle \nabla\Phi, \nabla\Phi \rangle_p := \frac{1}{2} \sum_{\{j,k\} \in E} (\Phi_j - \Phi_k)^2 g_{j,k}(p). \tag{4.1.3}$$

The above inner product provides the following distance on $\mathscr{P}_o(S)$.

Definition 4.1.3. *Given two discrete probability measures p^0, $p^1 \in \mathscr{P}_o(S)$, the Wasserstein metric $\mathcal{W}$ is defined by:*

$$\mathcal{W}(p^0, p^1) := \inf \left\{ \int_0^1 \langle \nabla\Phi, \nabla\Phi \rangle_p dt; \quad \frac{d}{dt} p = div(p\nabla\Phi) = 0, \quad p(0) = p^0, \quad p(1) = p^1, \quad p \in \mathcal{C}_t(\mathscr{P}(\mathbf{S})) \right\} \tag{4.1.4}$$

where $(p\nabla\Phi)_{j,k} = g_{j,k}(p)\nabla\phi_{j,k}$, for each $\{j,k\} \in E$.

As the solution for the Fokker-Planck equation it can be seen as conditioned probability, restriction $\frac{d}{dt}p = \text{div}(p\nabla\Phi) = 0$, implies that $p(t)$ is a probability density for each $t \in [0, 1]$.

Now we state one of the main theorems of this thesis, this result is due to Shue Nee, et al [18].

Theorem 4.1.4 (Evolutionary Dynamics)**.** *Given a potential game with strategy graph $G = (S, E)$, potential $\mathcal{F}(p) \in \mathbf{C}^2(\mathbb{R}^n)$ and a constant $\beta \geq 0$. Then the gradient flow of free energy*

$$-\mathcal{F}(p) + \beta \sum_{j=1}^{n} p_j \log p_i \tag{4.1.5}$$

on the Riemannian manifold $(\mathscr{P}_o(S), \mathcal{W})$ *is the Fokker-Planck equation*

$$\frac{d}{dt}p_k = \sum_{j\in N(k)} p_j \max\{F_k(p) - F_j(p) + \beta(\log p_j - \log p_k), 0\}$$
$$- \sum_{j\in N(k)} p_k \max\{F_j(p) - F_k(p) + \beta(\log p_k - \log p_j), 0\}$$

for any $k \in S$*. In addition, for any initial* $p^0 \in \mathscr{P}_o(S)$*, there exist a unique solution* $p(t) : [0, \infty) \to \mathscr{P}_o(S)$*.*

Is worth to remark that $\frac{d}{dt}p_k$ represents how fast the proportion of the population is changing to strategy $k \in S$. On the one hand, if $\frac{d}{dt}p_k$ is negative, several agents are leaving strategy k. By the other hand f $\frac{d}{dt}p_k$ is positive, several agents are taking strategy k. Thus, this intensity in the switching strategies, is modelling the combination of taking risk of the agents and their search of wellness. Some interesting question are if the solution of the Fokker Plank equation may converge to a Nash equilibrium. Indeed it does and is related to fixed point for (4.1.5) and Gibbs measures. we refer to the reader for the proof of 4.1.4 and for more information about the relation of Gibbs measures and population games to [52].

Concluding remarks

The results presented in this Chapter have a more geometrical perspective. In particular the proof of 4.1.4– which we refer the reader for it to [52]– relies on properties of Riemmanian manifolds about how the gradient flow is completely determined by the vector space of payoff functions $(F_k)_{k\in S}$. For a general introduction of differential geometry approach in optimal transport we refer to [49, Part II] and [1]

Chapter 5

Conclusions

In this thesis we have briefly introduced several problems related to optimal. Thus. optimal transport seems to be a useful tool in different areas of application. In specific, we recall that Theorems 1.1.4 and 1.3.6 are of importance. Nevertheless, the calculation of an optimal transference plan is not easy at all. Several works have appeared to find explicitly those transference plans. For example, we refer to [47, Chapter 2], where –under the assumption that the cost function is of the form $c(x,y) = h(\|x-y\|)$, where h is some differentiable convex function– the optimal tranference plan is deterministic and is given by a change of variables formula of other convex function. Another relevant approach was introduced by C. Villani [49, Part II], where he introduces several notions of differential geometry and using some variational approaches for finding those plans. Even so, under more general assumptions no work have been made or are still emerging. We can finally deduce that the universality of the optimal transport may be of interest as long as more works about finding or constructing those optimal transference plans arise in literature.

However optimal transport has established a break-point for probability theory and is just the the beginning of a number of optimisations problems; some emerging problems to take on consideration could be:

- Consider a family of probability spaces, $(\mathbf{X}_j, \mathcal{F}_j, \mu_j)_{j=1}^N$ and a measurable cost function $c : \mathbf{X} := \mathbf{X}_1 \times \cdots \times \mathbf{X}_N \to \mathbb{R}$. The aim is to minimized the total cost

$$\int_{\mathbf{X}} c(x_1, \ldots, x_N) d\pi(x_1, \ldots, x_N)$$

  over the set of all probability measures on the measurable space $(\mathbf{X}, \bigotimes_{j=1}^M \mathcal{F}_j)$, such that π have marginals μ_j in the space $\mathbf{X}_j$, for all $j = 1, \ldots, N$. Future works may be orientated in this topic, and to its applications to optimal control. We refer to [41] for a general approach to this problem.

- Under the same assumption for the optimal transport problem (1.0.3), consider an additional condition for the coupling (X, Y) of (μ, ν), that $\mathbb{E}(X|Y) = X$. This condition is known as the martingale restriction. Thus, the infimum runs over all couplings holding the martingale restriction. We refer to [20, 5], for instance for a formal description to the problem.

- Finally, we can consider the notion of games with infinitely-many players, as in Chapters 3 and 5. In this case, however, the dynamics of each agent is ruled by some fixed optimal control problem. This games are known as mean field games, and they carry with them a system of two differential equations known in literature as the master equations; one is given by a backwards Hamilton-Jacobi equation, which describes the flow of the population through a vector field, and the other is a mass conservation equation, which describes how the population density evolves in time, with a given initial density and initial state. The main difficulty is that only rare problems admit classical solutions to the systems. Moreover, only "toy" examples admit explicit solutions for the system. Therefore, a lot of theory for viscosity and distributional solutions for the systems has been developed in the area. For more information about mean field games, we refer to [34, 11, 33]. Future works may be oriented to use the optimal transport approach, in order to solve those systems or to find numerical solutions for them, see [12] for a picture of the problematic.

We conclude with We fact that optimal transport shares several relation with game theory and optimal control. Furthermore, we have found approaches for finding solution for different problems, (e.g), the topological properties of optimal transport –Wasserstain distances– were crucial the analysis of convergence in Theorem 2.3.2

We are aware that more examples in this work may be of help at the time of understanding the main objects of study, and also for the main results in this thesis (see Theorems 2.2.8, 1.3.6, 2.2.4, 3.1.6, 3.2.1, and 4.1.4), but for instance we refer to [9, 26, 13, 48] for Chapter 2 examples, to [7, 14, 15, 8] for Chapter 4 examples, and to [18, 40, 16, 52] for Chapter 5 examples.

Appendix A

Appendix

Here we state some classical result about real and functional analysis [46, 42], some notion of [44], and some concepts of probability theory [6].

In this section $\mathcal{X}$ will denote a Banach space, $\mathcal{X}'$ its topological dual space, and $\mathbf{X}$ a Polish space.

Definition A.0.1. *A sequence $\{x_n\}_{n=0}^{\infty}$ in $\mathcal{X}$ is said to be weakly converging to x, which we denote $x_n \rightharpoonup x$, if we have $\langle \xi, x_n \rangle \to \langle \xi, x \rangle$, for every $\xi \in \mathcal{X}'$. A sequence $\{\xi_n\}_{n=0}^{\infty} \subset \mathcal{X}'$ is said to be weakly-*converging to ξ, and we denote $\xi_n \overset{*}{\rightharpoonup} \xi$ if $\langle \xi_n, x \rangle \to \langle \xi, x \rangle$ for every $x \in \mathcal{X}$.*

Theorem A.0.2. *If $\mathcal{X}$ is a separable Banach space, and $\{\xi_n\}_{n=0}^{\infty}$ is a bounded sequence in $\mathcal{X}'$, then there exists a subsequence $\{\xi_{n_k}\}_{k=0}^{\infty}$ weakly converging to some $\xi \in \mathcal{X}'$.*

Theorem A.0.3 (Riesz representation theorem)**.** *Suppose that $\mathbf{X}$ is a separable and locally compact metric space. Let $\mathcal{X} = \mathbf{C}_0(\mathbf{X})$ be the space of continuous functions on $\mathbf{X}$ vanishing at infinity, endowed with the supremum norm, which makes the space complete. As $\mathbf{C}_0(\mathbf{x})$ is a Banach space, it is a closed subset of $\mathbf{C}_b(\mathbf{x})$. Then every element of $\mathcal{X}'$ is represented in a unique way as an element of $\mathscr{M}(\mathbf{X})$; for each $\xi \in \mathcal{X}'$, there is a unique $\lambda \in \mathscr{M}(\mathbf{X})$ such that $\langle \xi, \phi \rangle = \int \phi d\lambda$; moreover, $\mathcal{X}'$ is isomorphic to $\mathscr{M}(\mathbf{X})$ endowed with the norm $\|\lambda\| := |\lambda|(\mathbf{X})$ of total variation.*

We denote $\mathscr{P}(\mathbf{X})$ as the space of probability measures on $\mathbf{X}$. We restrict ourselves to Borel probability measures.

For the proof to the next theorem we refer to [47, Chapter 2]

Theorem A.0.4. *Let $f : \mathbf{X} \to \mathbb{R}$ be a function bounded from bellow. Then f is lower semi-continuous if and only if there exist a sequence f_k of bounded Lipschitz functions such that for every $x \in \mathbf{X}$, $f_k(x)$ converges increasingly to $f(x)$.*

Theorem A.0.5 (Weierstrass)**.** *If $f : \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ is lower semi-continuous and $\mathbf{X}$ is compact, then there exist $\overline{x} \in \mathbf{X}$ such that $f(\overline{x}) = \min\{f(x) : x \in \mathbf{X}\}$*

Definition A.0.6. *We say that $\{\mu_n\}_{n=0}^{\infty} \subset \mathscr{P}(\mathbf{X})$ weakly converges to μ if, for each $\phi \in \mathbf{C}_b(\mathbf{X})$*

$$\lim_{n\to\infty} \int_{\mathbf{X}} \phi d\mu_n = \int_{\mathbf{X}} \phi d\mu,$$

and we write $\mu_n \rightharpoonup \mu$.

Remark. *Note that when* $\mathbf{X}$ *is a compact space, the weak and weak* convergence are the same.*

Theorem A.0.7. *Let* μ *and* ν *in* $\mathscr{P}(\mathbf{X})$ *such that, for each* $\phi \in \mathbf{C}_b(\mathbf{X})$,

$$\int_{\mathbf{X}} \phi d\mu = \int_{\mathbf{X}} \phi d\nu.$$

Then $\mu \equiv \nu$.

Theorem A.0.8. *Let* $\{\mu_n\}_{n=0}^{\infty}, \mu$ *in* $\mathscr{P}(\mathbf{X})$, *such that,* $\mu_n \rightharpoonup \mu$, *and* $T : \mathbf{X} \to \mathbb{R}$ *a* $\mu-$*almost surely continuous function. Then* $T\#\mu_n \rightharpoonup T\#\mu$.

The following theorem gives us a "characterisation" of compactness, based on the concept of tightness (see definition 1.0.6), for any family of probability measures, and will be essential for the structures that will be developed in next chapters. The proof can be consulted in [6].

Theorem A.0.9 (Prokhorov)**.** *Suppose that* $\{\mu_n\}_{n=0}^{\infty} \subset \mathscr{P}(\mathbf{X})$ *is tight over a Polish space* $\mathbf{X}$. *Then there exists* $\mu \in \mathscr{P}(\mathbf{X})$ *and a subsequence* $\{\mu_{n_k}\}_{k=0}^{\infty} \subset \{\mu_n\}_{n=0}^{\infty}$ *such that* $\mu_{n_k} \rightharpoonup \mu$, *(in duality with* $\mathcal{C}_b(\mathbf{X})$*). Conversely, every sequence* $\{\mu_n\}_{n=0}^{\infty} \subset \mathscr{P}(\mathbf{X})$ *weakly converging to some* $\mu \in \mathscr{P}(\mathbf{X})$ *is tight.*

The next theorem is a well known result from functional analysis, and for the proof we refer to [42], but as an almost straightforward observation, theorem $A.0.11$ follows and give us a way to construct measures with compact support from limit of measures with discrete support.

Theorem A.0.10. *Consider the Banach algebra* $C(X)$, *with* X *a compact Hausdorff space, equipped with the pointwise algebraic operation and the supremum norm. Let* $M(X)$ *represent the Dual space of* $C(X)$ *and* $P(X)$ *denote the subset consisting of those* μ *in* $M(X)$ *such that* $\|\mu\|_{TV} = 1$. *Then* $P(X)$ *is a convex, weakly-*$*$ *compact set, whose extremal points are the Dirac measures* δ_x, $x \in X$, *given by* $\delta_x f = f(x)$, *for every* $f \in C(X)$.

Theorem A.0.11. *Let* X *be a locally compact Hausdorff space. Then the set of elements in* $\mathscr{P}(X)$ *with support contained in a compact set* $M \subset X$, *is a convex and weakly-*$*$ *compact set. Additionally, the set of discrete measures with support contained in* M *are weakly-*$*$ *dense in the subset of* $\mathscr{P}(X)$ *with support in* M.

The next observation and the next result are taken from the book [21, Chapter 11]. This remarks are a generalisation of the convergence of empirical measures on general metric spaces.
Consider the Polish probability space, $(\mathbf{X}, \mathcal{F}, \mu)$. Then there exists a probability space $(\mathbf{\Omega}, \mathbb{P})$ on which there are independent random variables $\{X_j\}_{j=1}^{\infty}$ with values in $\mathbf{X}$ and law $(X_j) = \mu$ for every $j \in \mathbb{N}$. Since we can take $\mathbf{\Omega}$ as a Cartesian product of a sequence of copies of $\mathbf{X}$ and X_j as coordinates. For $n \in \mathbb{N}$, the empirical measure μ_n is defined by

$$\mu_n(A)(\omega) := \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j(\omega)}(A),$$

for each $A \in \mathcal{F}$ and $\omega \in \mathbf{\Omega}$.

Theorem A.0.12 (Varadarajan)**.** *Let* $(\mathbf{X}, \mathcal{F})$ *be a Polish measurable space and* μ *any Borel probability measure on* $(\mathbf{X}, \mathcal{F})$*. Then, the empirical measures* $\{\mu_n\}_{n=1}^{\infty}$ *converge to* μ*,* $\omega-$*almost surely:*

$$\mathbb{P}(\{\omega \in \mathbf{\Omega} : \mu_n(\cdot)(\omega) \rightharpoonup \mu\}) = 1.$$

Bibliography

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] R. J. Aumann. Markets with a continuum of traders. *Econometrica*, 32:39–50, 1964.

[3] R. J. Aumann. Existence of competitive equilibria in markets with a continuum of traders. *Econometrica*, 34:1–17, 1966.

[4] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations.* Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 1997. With appendices by Maurizio Falcone and Pierpaolo Soravia.

[5] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016.

[6] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons, Inc., New York, second edition, 1999.

[7] A. Blanchet and G. Carlier. From Nash to Cournot-Nash equilibria via the Monge-Kantorovich problem. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20130398, 11, 2014.

[8] A. Blanchet and G. Carlier. Optimal transport and Cournot-Nash equilibria. *Math. Oper. Res.*, 41(1):125–145, 2016.

[9] J. Blanchet, H. Lam, Q. Tang, and Z. Yuan. Robust actuarial risk analysis. *N. Am. Actuar. J.*, 23(1):33–63, 2019.

[10] G. Buttazzo, A. Pratelli, S. Solimini, and E. Stepanov. *Optimal urban networks via mass transportation.* Springer-Verlag, Berlin, 2009.

[11] P. Cardaliaguet. Mean field games: the master equation and the mean field limit. In *Séminaire Laurent Schwartz—Équations aux dérivées partielles et applications. Année 2015–2016*, pages Exp. No. XVII, 10. Ed. Éc. Polytech., Palaiseau, 2017.

[12] P. Cardaliaguet, A. R. Mészáros, and F. Santambrogio. First order mean field games with density constraints: pressure equals price. *SIAM J. Control Optim.*, 54(5):2672–2709, 2016.

[13] G. Carlier and I. Ekeland. Equilibrium structure of a bidimensional asymmetric city. *Nonlinear Anal. Real World Appl.*, 8(3):725–748, 2007.

[14] L. Chen, T. Ho, M. Chiang, S. H. Low, and J. C. Doyle. Congestion control for multicast flows with network coding. *IEEE Trans. Inform. Theory*, 58(9):5908–5921, 2012.

[15] M.-W. Cheung and R. Lahkar. Nonatomic potential games: the continuous strategy case. *Games Econom. Behav.*, 108:341–362, 2018.

[16] S.-N. Chow, L. Dieci, W. Li, and H. Zhou. Entropy dissipation semi-discretization schemes for fokker–planck equations. *J. Dynam. Differential Equations*, 31(2):765–792, 2019.

[17] S.-N. Chow, W. Li, J. Lu, and H. Zhou. Population games and discrete optimal transport. *J. Nonlinear Sci.*, 29(3):871–896, 2019.

[18] S.-N. Chow, W. Li, and H. Zhou. Entropy dissipation of Fokker-Planck equations on graphs. *Discrete Contin. Dyn. Syst.*, 38(10):4929–4950, 2018.

[19] C. Dellacherie. Ensembles analytiques. Théorèmes de séparation et applications. pages 336–372. Lecture Notes in Math., Vol. 465, 1975.

[20] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields*, 160(1-2):391–427, 2014.

[21] R. M. Dudley. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2002.

[22] I. Ekeland. Notes on optimal transportation. *Econom. Theory*, 42(2):437–459, 2010.

[23] K. Elamvazhuthi, P. Grover, and S. Berman. Optimal transport over deterministic discrete-time nonlinear systems using stochastic feedback laws. *CoRR*, abs/1804.03108, 2018.

[24] L. R. Ford, Jr. and D. R. Fulkerson. *Flows in networks*. Princeton University Press, Princeton, N.J., 1962.

[25] A. Galichon. *Optimal transport methods in economics*. Princeton University Press, Princeton, NJ, 2016.

[26] A. Galichon. A survey of some recent applications of optimal transport methods to econometrics. *Econom. J.*, 20(2):C1–C11, 2017.

[27] C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Math. J.*, 31(2):231–240, 1984.

[28] J. González-Hernández, J. R. Gabriel, and O. Hernández-Lerma. On solutions to the mass transfer problem. *SIAM J. Optim.*, 17(2):485–499, 2006.

[29] O. Hernández-Lerma and J. B. Lasserre. *Discrete-time Markov control processes*, volume 30 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1996. Basic optimality criteria.

[30] O. Hernández-Lerma and J. B. Lasserre. *Further topics on discrete-time Markov control processes*. Springer-Verlag, New York, 1999.

[31] L. Kantorovitch. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[32] S. S. Keerthi and E. G. Gilbert. Optimal infinite-horizon feedback laws for a general class of constrained discrete-time systems: stability and moving-horizon approximations. *J. Optim. Theory Appl.*, 57(2):265–293, 1988.

[33] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris*, 343(9):619–625, 2006.

[34] J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.

[35] J.-L. Lions. *Optimal control of systems governed by partial differential equations*. Springer-Verlag, New York-Berlin, 1971.

[36] D. Monderer and L. S. Shapley. Potential games. *Games Econom. Behav.*, 14(1):124–143, 1996.

[37] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale, 1781.

[38] J. Nash. Non-cooperative games. *Ann. of Math. (2)*, 54:286–295, 1951.

[39] J. F. Nash, Jr. Equilibrium points in n-person games. *Proc. Nat. Acad. Sci. U. S. A.*, 36:48–49, 1950.

[40] M. A. Nowak. *Evolutionary dynamics*. The Belknap Press of Harvard University Press, Cambridge, MA, 2006.

[41] B. Pass. Multi-marginal optimal transport and multi-agent matching problems: uniqueness and structure of solutions. *Discrete Contin. Dyn. Syst.*, 34(4):1623–1639, 2014.

[42] G. K. Pedersen. *Analysis now*. Springer-Verlag, New York, 1989.

[43] L. Rifford. *Sub-Riemannian geometry and optimal transport*. SpringerBriefs in Mathematics. Springer, Cham, 2014.

[44] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer-Verlag, Berlin, 1998.

[45] I. Rodnianski and B. Schlein. Quantum fluctuations and rate of convergence towards mean field dynamics. *Comm. Math. Phys.*, 291(1):31–61, 2009.

[46] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, third edition, 1987.

[47] F. Santambrogio. *Optimal transport for applied mathematicians.* Birkhäuser/Springer, Cham, 2015.

[48] L. Vanneschi, W. S. Bush, and M. Giacobini, editors. *Evolutionary computation, machine learning and data mining in bioinformatics.* Springer, Heidelberg, 2013.

[49] C. Villani. *Optimal transport. Old and.* Springer-Verlag, Berlin, 2009.

[50] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, New Jersey, 1944.

[51] Q. Xia. Motivations, ideas and applications of ramified optimal transportation. *ESAIM Math. Model. Numer. Anal.*, 49(6):1791–1832, 2015.

[52] H. Zhang, W. L. Li, Z. Z. Li, and Z. H. Ma. Evolutionary dynamics on regular graphs. *J. Lanzhou Univ. Nat. Sci.*, 45(6):121–124, 2009.