Bachelor's degree thesis

Analysis and design of Multi-agent Coverage and Transport algorithms

Melcior Pijoan Comas

Advised by:
Sonia Martínez (UCSD)
Arnau Dòria Cerezo (UPC)

In partial fulfillment of the requirements for the
Bachelor's degree in Mathematics
Bachelor's degree in Engineering Physics

May 2020

Abstract

Multi-agent robotic systems have shown to be useful and reliable solutions to many problems that arise in science and engineering. In this work we will study Coverage Control, that aims to achieve optimal coverage of a density. We will focus on the case when the density has a time dependence and we will study a Singular Perturbation Theory approach to solve the problem. We will also consider large swarms of agents, where we can develop continuous models to analyze the behaviour of the swarm. Recent work has focused on applying ideas from the theory of Optimal Transport to the Multi-Agent Transport problem. We will review the work and provide some modifications.

Keywords: Multi-agent systems, Voronoi Partitions, Coverage Control, Singular Perturbation Theory, Optimal Transport, Distributed online optimization, iterative scheme, proximal point algorithm

AMS Code: 35B35, 35B40, 49M25, 90C46, 93D05, 93D20

Acknowledgements

I would like to thank Professor Sonia Martínez for accepting me in her research group, being patient when my research progressed slowly, letting me join the reading group and teaching me how to be a competent researcher. I am also grateful to Vishaal Krishnaan for his help and collaboration in my project and his good disposition to always help me. I also want to thank Arnau Dòria, my UPC tutor.

With this thesis I finish my five year double degree in Mathematics and Engineering Physics and I am grateful to CFIS and Fundació Cellex for making my studies possible and providing a financial aid that allowed me to study abroad.

During these years I have made great friends, who are amazing people and have always made the studies much more fun.

I also believe that these last six months in San Diego have been an amazing experience and I am grateful to the people that I have met for making it a great experience.

Finally, I am grateful to my family for the support they give me in pursuing my studies and their financial aid with my studies.

Melcior Pijoan Comas

Santa Cristina d'Aro, May 2020

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Multi-agent systems are systems formed by multiple agents that can sense the environment, communicate with their neighbours and perform computations. Their characteristics present many uses in science and engineering, like the monitorization of environmental processes [17], search and rescue operations in hazardous environments [47] or joint actuation [34].
This motivates us to develop good algorithms that will allow the agents to perform the task efficiently and usually in a decentralized manner, since central computations require a higher communications cost and there might be situations in which the agents loose contact with the planner, for example exploring complex environments like caves or the ocean.
There are many interesting problems that appear in Multi-agent systems but we will focus on two problems, Coverage Control and large-scale Transportation. Coverage control [13], [12], [11] is a problem that appears naturally when we only have a limited amount of agents and we want them to achieve the optimal configuration in order to detect or track a density function. Coverage control can be used for many task such as environmental monitoring and search and rescue operations.
Our discussion of Coverage Control will be particularly focused on the coverage of time-varying densities, a problem that was discussed in [12], [31], [18] and [25]. We have participated in the work introduced in [25] where the theory of Singular Perturbations was used to develop a new approach to Coverage Control. We have provided additional experimental results and minor changes in the theory. We hope to publish our results soon.
The other application that we will study is Transport, which appears in many related problems such as extremum seeking of a potential field with chemotaxis [39], coverage control [13], [12], [11] and formation control [38]. We will particularly focus on large-scale swarms, which can be composed of miniature robots and have a lot of potential applications in biology [24], and engineering [46].
There have been different approaches in the literature to treat this problem, some of them include Markov Transition matrices [5], [7], [16] and continuum models [29], [22]. We will work with continuum models, in which the swarm can be abstracted as a fluid and we can study the evolution of the swarm by studying macroscopic variables like the density. We will develop control laws that act locally on each agent and we will study the macroscopic behaviour of the swarm. Studying the swarm macroscopically will allow us to use theoretical tools such as Lyapunov analysis and PDE analysis in order to prove convergence of the algorithms.

In order to study the transport of large scale Multi-Agent systems we will introduce Optimal Transport, a theory that deals with the optimal transportation or rearrangement of probability measures. The ideas of optimal transport promise an efficient way of transporting the swarm of agents and there have been some approaches [6] [14], [30] to introduce them in Multi-Agent swarms. Our research has focused in the recent work of [30], that provides a decentralized Optimal Transport algorithm for Multi-Agent Transport. Our contributions are twofold, we have improved the convergence rate of the algorithm and we have developed simulations in domains that present obstacles and therefore are non-convex.

Our work is structured as follows, in Chapter 2 we will introduce Coverage Control, formulate it as an allocation problem and provide a solution using Lloyd's algorithm following [13] and [12]. We will then focus on the coverage of time-varying densities, showing that Lloyd algorithm may not converge and providing and alternative algorithm [31] [18], which can be implemented in a centralized way or a decentralized way. Finally we will focus on a Singular Perturbation Theory approach to Time Varying Coverage Control [25]. We have joined the project and we have contributed providing additional experimental validation and minor changes in the theory. We hope to publish our results soon.

We will continue by introducing Optimal Transport theory in Chapter 3, following [49], [50] and [51]. Optimal transport will give us a framework to study distances between probability measures and rearrange probability measures. Additionally we will study gradient flows, which will be fundamental in order to formulate our multi-agent transport algorithms.

In Chapter 4 we will study the relation between Optimal Transport and Coverage Control. We will consider the problem of Optimal Transport when one of the probability measures is discrete and the other is continuous following [50]. This will allow us to compare the costs of the two problems and see that the locational cost of Coverage Control is a relaxation of the Optimal Transport cost. We will also follow [28] to explore the continuous limit of Coverage Control, we will see that the Coverage Control problem is not well posed to be studied macroscopically.

Finally, we will present a Multi-Agent transport algorithm that uses Optimal Transport in Chapter 5. Most of the work of section comes from [30]. We have provided some modifications, improving the rate of convergence of the algorithm and performing additional simulations with an obstacle environment. We have also proposed different models to include collision avoidance to the algorithm but our approaches are still incomplete. We discuss our ideas to include collision avoidance but we will leave their implementation for future work.

Chapter 2

Coverage Control

Coverage Control is an optimal resource allocation problem [19], [52], [42], in which we want to allocate a group of agents in a configuration that minimizes the cost and provides the best quality-of service.
We will consider a mobile sensing network composed of agents that can sense their environment and communicate. We will define a distribution in the space that will represent a measure of the probability that some event happens and we will update the positions of the agents in order to find the configuration that minimizes the sensing cost.
Our work will be structured as follows, we will start introducing some concepts from Computational Geometry, which will help us solve Coverage Control problem. Then we will follow [13], [12] to define the Coverage problem and solve it using Lloyd's algorithm. We will then study the Coverage Control problem when the target distribution has a time dependence. In this situation Lloyd algorithm may not converge and there is a need to develop efficient algorithms that can provide a good solution to the problem. We will review the approaches of [31], [18] and we will present a Singular Perturbation Theory approach that was originally presented in [25]. We have collaborated in the project doing minor modifications in the theory and performing some additional simulations. We hope to publish our results soon.

2.1 Voronoi diagrams

We start introducing some notions of Computational Geometry, in particular Voronoi Diagrams. A detailed treatment on Voronoi Diagrams can be found in [15] and [41].
Given $\Omega \subset \mathbb{R}^N$ we say that the sets $\mathcal{W} = \{W_1, \ldots, W_n\}$ are a partition Ω, if $\cup_i W_i = \Omega$ and their interiors are disjoint, $\mathring{W}_i \cap \mathring{W}_j = \emptyset$. Now we define the Voronoi partition of a set Ω given by some points p.

Definition 1 (Voronoi Partition). *Given n points in $\Omega \subset \mathbb{R}^N$, $P = (p_1, \ldots, p_n)$ with $p_i \in \Omega$, and a metric $c : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ we define the Voronoi Partition $\mathcal{V}(P) = (V_1, \ldots, V_n)$ generated by the points P as*

$$V_i = \{q \in \Omega | \, c(p_i, q) \leq c(p_j, q) \quad \forall j \neq i\}$$

We call V_i a Voronoi cell or Voronoi regions. When Ω is a convex set, the Voronoi cells will also be convex sets and the boundary of the Voronoi regions will be an hyperplane. In Figure

2.1 we illustrate the Voronoi Diagram, where we can see that the Voronoi cells assign every point $q \in \Omega$ to the closest p_i. Now we define a similar partition, the Voronoi power cell, which assigns points $q \in \Omega$ to points p_i according to the weight or bias w_i of the points.

Definition 2 (Voronoi Power Diagram)**.** *Given n points in* $\Omega \subset \mathbb{R}^N$*,* $P = (p_1, \ldots, p_n)$ *with* $p_i \in \Omega$*, a vector* $w = (w_1, \ldots, w_n)$*, and a metric* $c : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ *we define the Voronoi Power Diagram* $\mathcal{V}_w(P) = (V_{w,1}, \ldots, V_{w,n})$ *generated by the points P as*

$$V_i = \{q \in \Omega | \quad c(p_i, q) + w_i \leq c(p_j, q) + w_j \quad \forall i \neq j\}$$

If Ω is convex the Voronoi Power Cells are convex. In Figure 2.1 we compare the Voronoi diagram to the Voronoi Power Diagram.

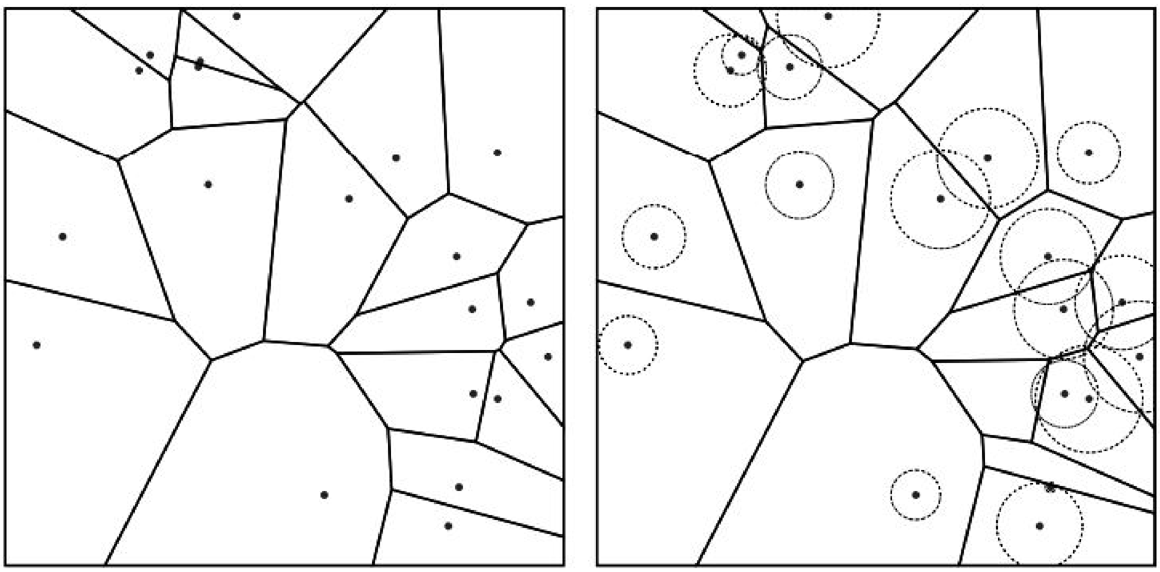

Figure 2.1: Comparison between the Voronoi Diagram (left) and the Voronoi Power Diagram (right), where the $w_i = r_i^2$, with r_i the radius of the circles centered around the points. Note that in the power diagram the generator may lie outside of the corresponding region. Figure extracted from [10]

2.2 Problem definition

We will now follow [13] and [12] to develop the theory of the Coverage Control Problem. Coverage Control is a problem that arises when searching for the optimal allocation of n sensors $P = (p_1, \ldots, p_n)$ in a space Ω in order to maximize the sensing of a density $\rho : \Omega \to \mathbb{R}_{\geq 0}$ which will represent a measure of the information that some event happens. Given the position of the sensors $P = (p_1, \ldots, p_n)$ and a partition $\mathcal{W} = \{W_1, \ldots, W_n\}$ of Ω we define the coverage cost as

$$\mathcal{H}(P, \mathcal{W}) = \sum_{i=1}^{n} \int_{W_i} ||x - p_i||^2 \rho(x) dx \tag{2.1}$$

In this allocation cost an agent i, with location p_i senses the density ρ in the cell W_i with a quadratic cost $||x-p_i||^2$, which will penalize more the points q that are farther away from p_i. We would like to optimize the coverage cost $\mathcal{H}(P, \mathcal{W}(P))$ with respect to the positions of the agents P and the choice of partition $\mathcal{W}(P)$. We start by choosing the optimal partition $\mathcal{W}(P))$. We remember the definition of Voronoi cell in (1)

$$V_i = \{q \in \Omega | ||x-p_i|| \leq ||x-p_j|| \quad \forall i \neq j\}$$

this gives us

$$||x-p_i|| \leq ||x-p_j|| \quad \forall x \in V_i, \forall j \neq i$$

If we take $W_i = V_i$ we can write the coverage cost as

$$\mathcal{H}(P) = \sum_{i=1}^{n} \int_{V_i} ||x-p_i||^2 \rho(x) dx = \int_{\Omega} \min_i ||x-p_i||^2 \rho(x) dx \tag{2.2}$$

And the Voronoi partition is the optimal partition.

2.2.1 Lloyd's algorithm

Now we show how we can optimize the coverage cost (2.1) with respect to the position of the agents. We will use Lloyd's algorithm, which was originally presented in Quantization Theory [33], a theory that deals with the efficient quantization of continuous signals into a discrete set of samples. Lloyd's algorithm can be formulated as a gradient descend algorithm in continuous time. We start by calculating

$$\frac{\delta \mathcal{H}}{\delta p_i}(P) = \frac{\delta \mathcal{H}}{\delta p_i}(P, \mathcal{V}(P)) = \int_{V_i} \frac{\delta}{\delta p_i} ||x-p_i||^2 \rho(x) dx$$

We define the mass and the centroid of the Voronoi cells V_i

$$m_i = \int_{V_i} \rho(x) dx \quad c_i = \frac{1}{m_i} \int_{V_i} x \rho(x) dx$$

It can then be proved [13] that

$$\frac{\delta \mathcal{H}}{\delta p_i}(P) = 2 m_i (p_i - c_i) \tag{2.3}$$

From the derivative of the coverage cost we can deduce that we will have a minimum if $p_i = c_i$. We will call this configurations Centroidal Voronoi Tesselations. A detailed study on Centroidal Voronoi Tesselations can be found on [21]. We particularly remark the following property.

Remark 1. *There are no guarantees that there exists a unique minimum for the problem. An arbitrary pair Ω, ρ admits in general multiple centroidal Voronoi configurations.*

Due to Remark 1 we can only guarantee that our algorithm converges to one of the local minima. Now we present the continuous time implementation of Lloyd's algorithm. We consider the dynamics

$$\dot{p}_i = u_i$$

With

$$u_i = -k(p_i - c_i) \tag{2.4}$$

This dynamics is known as Lloyd descend and we can prove it's convergence to a Centroidal Voronoi Tesselation.

Proposition 1. *For closed loop system given by (2.4) the sensors location converges asymptotically to the set of critical points of $\mathcal{H}(P)$.*

Proof. The control law gives us

$$\frac{d\mathcal{H}}{dt}(P(t)) = \sum_{i=1}^{n} \frac{d\mathcal{H}}{dp_i}\dot{p}_i = -2k\sum_{i=1}^{n} m_i||p_i - c_i||^2$$

And the dynamics converge to $\mathcal{H}^{-1}(0)$, which is the set of Centroidal Voronoi Configurations. □

2.3 Time-varying Coverage problem

Our next objective will be studying the Coverage problem when the density $\rho(x,t)$ has a time dependence, in this situation Lloyd's algorithm may not converge to a Centroidal Voronoi Tesselation. In this section we follow the work of [31] and [18] to study algorithms that are better suited for the task of coverage in a time-varying setting. We start by showing the problems that arise with Lloyd's algorithm. In continuous time Lloyd's update is given by

$$\dot{p}_i(t) = -k(p_i(t) - c_i(p(t),t))$$

and the coverage cost evolves as

$$\frac{d\mathcal{H}}{dt}(P(t),t) = \frac{\delta\mathcal{H}}{\delta t} + \sum_{i=1}^{n} \frac{\delta\mathcal{H}}{\delta p_i}\dot{p}_i =$$

$$\sum_{i=1}^{n}\int_{V_i} ||x - p_i(t)||^2 \frac{\delta\rho}{\delta t}(x,t)dx - 2k\sum_{i=1}^{n} m_i||p_i(t) - c_i(p(t),t)||^2$$

In the last expression, if $\frac{\delta\rho}{\delta t}$ is big there are no guarantees that $\frac{d\mathcal{H}}{dt}$ will decrease. To solve this problem in [31] they propose a control law composed of two parts, convergence to a Centroidal Voronoi configuration and tracking of the Centroidal Voronoi configuration. We start by assuming

$$p_i(t) = c_i(t)$$

Differentiating the expression,

$$\dot{p}_i = \frac{dc_i}{dt} = \frac{\delta c_i}{\delta t} + \frac{\delta c_i}{\delta p}\dot{p}_i$$

Giving us the tracking control law

$$\dot{p} = \left(I - \frac{\delta c}{\delta p}\right)^{-1} \frac{\delta c}{\delta t} \tag{2.5}$$

In order to ensure convergence to a Centroidal Voronoi configuration we can add a Lloyd descend term.

$$\dot{p} = \left(I - \frac{\delta c}{\delta p}\right)^{-1} \left(-k(p(t) - c(p(t), t)) + \frac{\delta c}{\delta t}\right) \tag{2.6}$$

This algorithm is known as TVD-C, which stands for Time Varying Densities Centralized case and it can be proved to converge to a Centroidal Voronoi configuration.

Proposition 2. *If* $1 \notin eig(\frac{\delta c}{\delta p})$ *and we let*

$$\dot{p} = \left(I - \frac{\delta c}{\delta p}\right)^{-1} \left(-k(p(t) - c(p(t), t)) + \frac{\delta c}{\delta t}\right)$$

then $||p(t) - c(p(t), t)|| \to 0$ *as* $t \to +\infty$ *with a rate of decrease given by* $e^{(-k(t-t_0))}$

Proof. We consider the evolution of $||p(t) - c(p(t), t)||^2$,

$$\frac{d}{dt}(||p(t) - c(p(t), t)||^2) = 2(p - c)^T(\dot{p} - \dot{c}) =$$

Expanding $\dot{c}$,

$$= 2(p - c)^T \left(\left(I - \frac{\delta c}{\delta p}\right)\dot{p}(t) - \frac{\delta c}{\delta t}\right) =$$

Substituting the control law

$$= 2(p - c)^T \left(\left(I - \frac{\delta c}{\delta p}\right)\left(I - \frac{\delta c}{\delta p}\right)^{-1}\left(-k(p - c) + \frac{\delta c}{\delta t}\right) - \frac{\delta c}{\delta t}\right) =$$

$$= 2(p - c)^T \left(\left(-k(p - c + \frac{\delta c}{\delta t}\right) - \frac{\delta c}{\delta t}\right) = -2k||p - c||^2$$

From this we get $||p - c||^2 \propto e^{-2k(t-t_0)}$ and $||p - c|| \propto e^{-k(t-t_0)}$ □

This algorithm presents two problems, the first problem is the existence of the inverse. In [20] they showed that if $\rho(x)$ is log-concave the inverse is well defined in a neighbourhood of a Centroidal Voronoi configuration but we cannot guarantee the existence of the inverse in more general situations.

The other problem is related to the centralized nature of this algorithm, making it unsuitable for coverage of large swarms. To understand this better we provide the formulas for the terms of $\frac{\delta c}{\delta p}$ and $\frac{\delta c}{\delta t}$, developed in [31].

$$\begin{aligned} \left[\frac{\delta c}{\delta p}\right]_{ij} &= \frac{\delta c_i}{\delta p_j} = -\frac{\int_{\delta V_{ij}} (x - c_i)(x - p_j)^T \rho dx}{m_i ||p_i - p_j||} \\ \left[\frac{\delta c}{\delta p}\right]_{ii} &= \frac{\delta c_i}{\delta p_i} = \sum_{j \in \mathcal{N}_i} \frac{\int_{\delta V_{ij}} (x - c_i)(x - p_i)^T \rho dx}{m_i ||p_i - p_j||} \\ \frac{dc_i}{dt} &= \frac{\int_{V_i} (x - c_i) \frac{\delta \rho}{\delta t} dx}{m_i} \end{aligned} \tag{2.7}$$

From the expression of $\frac{\delta c}{\delta p}$ we see that it has a sparse structure, being non-zero on the indexes corresponding to neighbouring agents. Despite this, the term $\left(I - \frac{\delta c}{\delta p}\right)^{-1}$ will be in general not sparse. To solve this problem in [31] they propose a Neumann series expansion,

$$\left(I - \frac{\delta c}{\delta p}\right)^{-1} \approx \sum_{l=0}^{k} \left(\frac{\delta c}{\delta p}\right)^{l} = I + \frac{\delta c}{\delta p} + \left(\frac{\delta c}{\delta p}\right)^{2} + \cdots + \left(\frac{\delta c}{\delta p}\right)^{l} \tag{2.8}$$

The constant k determines the order of the approximation and also the amount of neighbouring information needed, since $\left[\left(\frac{\delta c}{\delta p}\right)^{l}\right]_{ij} \neq 0 \implies j \in \mathcal{N}_l(i)$.

To ensure convergence of the Neumann series we require that $\lim_{l \to \infty} \left(\frac{\delta c}{\delta p}\right)^{l} = 0$ which corresponds to $|\lambda_{max}| < 1$. In [31] it was proven that the condition $|\lambda_{max}| < 1$ will be satisfied in a neighbourhood of a Centroidal Voronoi configuration, the author also remarked that the algorithm with $k \leq 1$ is always well defined and will always converge, while choosing a higher value of k may result in failure to converge.

The Neumann series approximation leads to the decentralized version of Time Varying Densities, TVD-D_k.

$$\dot{p} = \sum_{l=0}^{k} \left(\frac{\delta c}{\delta p}\right)^{l} \left(-k(p - c) + \frac{\delta c}{\delta t}\right) \tag{2.9}$$

2.4 Singular Perturbation Coverage

Another interesting approach to the Time-Varying Coverage problem was introduced in [25]. A Singular Perturbation Theory algorithm was proposed, the algorithm relied on a fast communication between agents in order to estimate the control law that updates the positions. The benefit of the algorithm is that it only needs nearest neighbour information and can recover the centralized solution if the communication speed is much higher than the position update speed. We have joined the project, making small changes in the theory and performing numerical experiments. We hope to publish our results soon.

2.4.1 Singular Perturbation Theory

In order to explain the algorithm we will first introduce Singular Perturbation Theory. Singular Perturbation Theory is particular case of Perturbation Theory, where the general approach is to obtain solutions as a power series of some small parameter ϵ. In the case of a Singular Perturbation the dynamics of the system cannot be approximated by setting the parameter $\epsilon = 0$. Singularly Perturbed methods are usually characterized by dynamics operating on multiple scales.
We consider the system

$$\begin{aligned} \dot{p} &= f((p,c),u,t) \\ \epsilon\dot{u} &= g((p,c),u,t) \end{aligned} \tag{2.10}$$

In this system we can find two different time scales, if take $\epsilon \to 0^+$ we find the reduced model

$$\dot{p} = f((p,c),u,t) \tag{2.11a}$$
$$0 = g((p,c),u,t) \tag{2.11b}$$

Here u is constant and we can find its explicit form as $u = h((p,c),t)$. Where h is a function of p, c, t which may not be uniquely defined. On the other hand, we can consider a stretched time-scale $\eta = t/\epsilon$ and analyze the dynamics of the boundary layer by taking $\epsilon \to 0^+$.

$$dp/d\eta = 0 \tag{2.12a}$$
$$du/d\eta = g((p,c),u,t) \tag{2.12b}$$

In the boundary layer p is frozen and doesn't evolve. To summarize, the reduced model (2.11) represents the dynamics at a slow time scale, neglecting the faster dynamics, and the boundary layer (2.12) represents the dynamics at a fast time scale neglecting the slower time-scale.
Finally, we will introduce a result from [27] which allows us to use the singular perturbation dynamics to approximate the the reduced model.

Theorem 1 (Tikhonov's theorem)**.** *If $u = h((p,c),t)$ is an exponentially stable root of the boundary layer dynamics (2.12b) the solution of the system (2.10) approaches the solution of the reduced model (2.11) as $\epsilon \to 0^+$. Note that there may be more than one such function h.*

2.4.2 Main result

Now we return to the Time-Varying Densities coverage problem. We want to build a system with two time scales, a fast communication time scale and a slow movement time scale. We start by considering the linear system

$$\left(I - \frac{\delta c}{\delta p}\right)\dot{p} = \left(-k(p-c)) + \frac{\delta c}{\delta t}\right) \tag{2.13}$$

If we solve the system inverting the matrix $\left(I - \frac{\delta c}{\delta p}\right)$ we will recover the TVD-C algorithm (2.6). Our objective is to solve this system of equations by minimizing the difference between the two sides of the equality. To do it we consider the function

$$F(u) = \left\| \left(I - \frac{\delta c}{\delta p} \right) u - \left(-k(p-c) + \frac{\delta c}{\delta t} \right) \right\|^2 \tag{2.14}$$

And we construct a singularly perturbed system,

$$\begin{aligned} \dot{p} &= u \\ \dot{u} &= -\frac{1}{\epsilon} \nabla F(u) \\ \text{with } \nabla F(u) &= \left(I - \frac{\delta c}{\delta p} \right)^T \left(\left(I - \frac{\delta c}{\delta p} \right) u - \left(-k(p-c) + \frac{\delta c}{\delta t} \right) \right) \end{aligned} \tag{2.15}$$

We will call this dynamics TVD-SP$_\epsilon$, from Time Varying Densities Singular Perturbation. In the next theorem we show it's convergence.

Theorem 2. *The dynamics of TVD-SP$_\epsilon$ (2.15) approach the dynamics given by TVD-C (2.6) as $\epsilon \to 0^+$. Additionally, for $0 < \epsilon \ll 1$, $||p - c||$ converges exponentially to zero.*

Proof. We start by studying the equations (2.15), with $\dot{p} = u$,

$$\dot{u} = -\frac{1}{\epsilon} \left(I - \frac{\delta c}{\delta p} \right)^T \left(\left(I - \frac{\delta c}{\delta p} \right) u - \left(-k(p-c) + \frac{\delta c}{\delta t} \right) \right)$$

The boundary layer will be given by

$$\frac{du}{d\eta} = -\left(I - \frac{\delta c}{\delta p} \right)^T \left(\left(I - \frac{\delta c}{\delta p} \right) u - \left(-k(p-c) + \frac{\delta c}{\delta t} \right) \right)$$

With $\eta = \frac{t}{\epsilon}$. This is a linear system and from the theory of linear systems we know that it has a unique solution and the solution is exponentially stable. This allows us to apply Theorem 1, and taking $\epsilon \to 0^+$ we recover the reduced model

$$\begin{aligned} \dot{p} &= u \\ 0 &= \left(I - \frac{\delta c}{\delta p} \right)^T \left(\left(I - \frac{\delta c}{\delta p} \right) u - \left(-k(p-c) + \frac{\delta c}{\delta t} \right) \right) \end{aligned} \tag{2.16}$$

Calculating the explicit value of u we find

$$\begin{aligned} \dot{p} &= u \\ u &= \left(I - \frac{\delta c}{\delta p} \right)^{-1} \left(-k(p-c) + \frac{\delta c}{\delta t} \right) \end{aligned} \tag{2.17}$$

Obtaining the dynamics of TVD-C (2.6). Finally, from the convergence we can extend the exponential decrease of $||p - c||$ in TVD-C to TVD-SP$_\epsilon$ for $0 < \epsilon \ll 1$. □

In order to implement (2.15), we must integrate the dynamics of p and u in their timescales.

The dynamics of p will be given by the reduced model of (2.15), which translates to

$$p[k+1] = p[k] + u\Delta t \tag{2.18}$$

Where u is considered constant. On the other hand, to discretize u we integrate the boundary layer of (2.15), giving us

$$u[l+1] = u[l] - \Delta\eta \left(I - \frac{\delta c}{\delta p}\right)^T \left(\left(I - \frac{\delta c}{\delta p}\right) u - \left(-k(p-c) + \frac{\delta c}{\delta t}\right)\right) \tag{2.19}$$

Where p is considered constant and $\eta = \frac{t}{\epsilon}$. The time steps Δt, $\Delta\eta$ must be chosen small enough to ensure convergence of their respective dynamics. Additionally for every update in p we have $N = \frac{\Delta t}{\epsilon \Delta\eta}$ updates in u with a frozen value of p. With this considerations, if ϵ is small enough the boundary layer dynamics will converge and we will recover the solution of TVD-C.

2.4.3 Experimental results

We will now present the the results of our simulations. We have implemented our algorithms in Python. We have compared the performance of the algorithms Lloyd, TVD-C, TVD-D$_k$, and TVD-SP$_\epsilon$ for various density functions using different numbers of agents. We have used the following time-varying densities as target:

$$\rho_1(x,t) = e^{-\left((x_1 - 2\sin(t/\tau))^2 + (x_2/4)^2\right)} \tag{2.20}$$

$$\rho_2(x,t) = e^{-\left((x_1 - \sin(t/\tau))^2 + (x_2 + \sin(2t/\tau))^2\right)} \tag{2.21}$$

$$\rho_3(x,t) = e^{-\left((x_1 - 2\cos(t/\tau))^2 + (x_2 - 2\sin(t/\tau))^2\right)} \tag{2.22}$$

Where $\tau = 5$. We measure the performance of the algorithms using the total cost

$$\int_0^T H(x,V,t)dt \tag{2.23}$$

In our experiments we have observed that since there are multiple Centroidal Voronoi Configurations we cannot guarantee that two different algorithms will converge to the same Centroidal Voronoi Configuration. This will introduce some uncertainty in the comparison of the different algorithms but we can still observe the general patterns.
We start comparing the cost of TVD-SP$_\epsilon$ with Lloyd and TVD-C with 10 agents. We have used $\Delta t = 5 \cdot 10^{-2}$ and $\Delta\eta = 10^{-4}$ to ensure convergence. In Figure 2.2 we can see how we can achieve a cost similar to TVD-C by taking a value of ϵ of the order of 10^{-2}. We can also see how if ϵ increases the cost can become higher and when $\epsilon > 10^{-1}$ we have encountered cases where the algorithm didn't converge and agents escaped the domain.

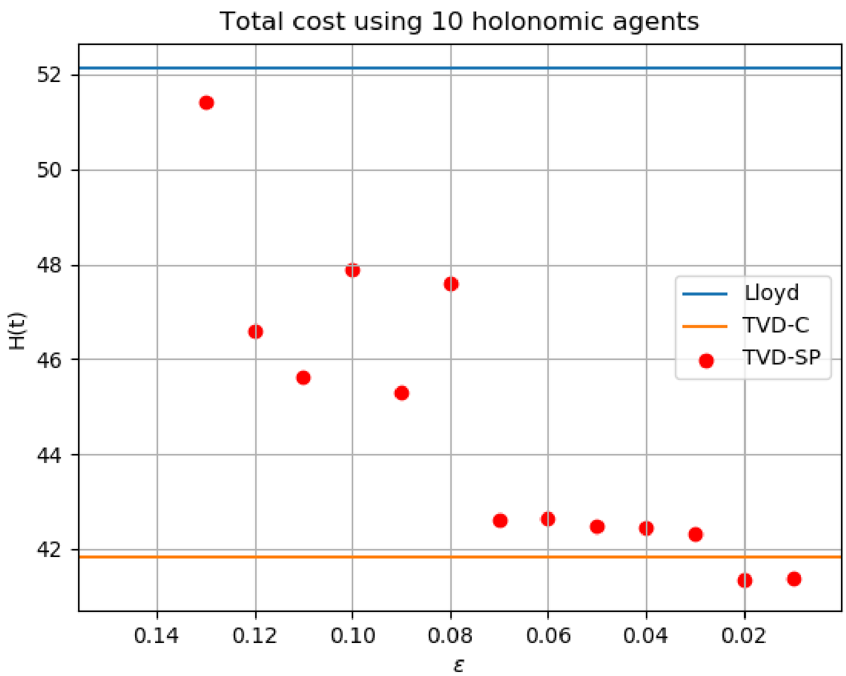

Figure 2.2: Comparison of TVD-SP with Lloyd and TVD-C for 10 agents

A possible solution to fix this problem is to reset the control law when the agents move. We propose to use

$$u[0] = -\kappa(p - c) + \frac{\delta c}{\delta t}$$

Which corresponds to the control law of TVD-D_0. We call this algorithm TVD-SSP_ϵ, for Time Varying Densities (modified) Start Singular Pertubation. This algorithm doesn't suffer from the convergence problems when ϵ is big because it acts as TVD-D_0 when ϵ is big. In Figure 2.3 we can see the comparative between the algorithm, TVD-C and Lloyd. The cost of TVD-SSP_ϵ converges much smoothly to the cost of TVD-C than TVD-SP_ϵ but it also requires a smaller value of $\epsilon = 10^{-3}$ to achieve a cost comparable to TVD-C.

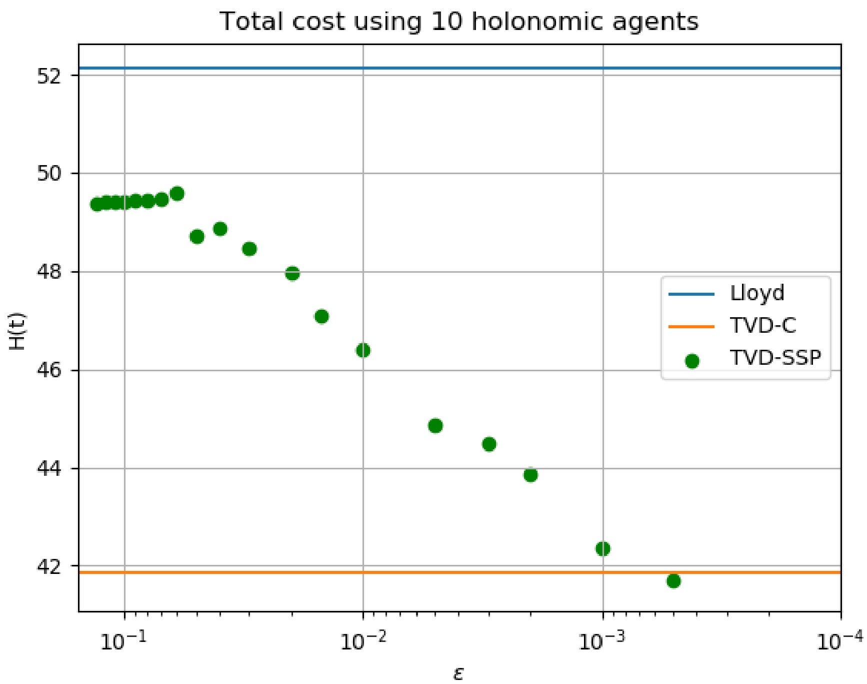

Figure 2.3: Comparison of TVD-SSP with Lloyd and TVD-C for 10 agents

Finally in Table 2.1 we have a comparative of all the algorithms with 10 agents. We note how by taking ϵ small enough with TVD-SP$_\epsilon$ and TVD-SSP$_\epsilon$ we can obtain better results than TVD-D$_k$ and achieve a similar cost as TVD-C.

Table 2.1: Comparison of all the algorithms with 10 agents

Algorithm	ρ_1	ρ_2	ρ_3
Lloyd	52.11	50.57	62.82
TVD-D$_0$	49.10	48.55	55.41
TVD-D$_1$	44.39	44.23	47.90
TVD-D$_2$	42.68	42.70	44.75
TVD-D$_3$	41.95	41.97	43.23
TVD-SP$_{0.1}$	46.26	46.26	53.29
TVD-SP$_{0.05}$	42.49	43.07	42.93
TVD-SP$_{0.01}$	41.40	41.45	41.48
TVD-SSP$_{0.01}$	46.38	48.67	49.68
TVD-SSP$_{0.005}$	44.87	45.94	46.61
TVD-SSP$_{0.001}$	42.37	41.97	42.97
TVD-C	41.84	42.47	41.46

After performing the experiment with 10 agents we would like to obtain the same results with a higher amount of agents. Unfortunately we have observed that as the number of agents increases some numerical problems appear. In particular we have very poor guarantees on conditioning of $\left(I - \frac{\delta c}{\delta p}\right)$, since the theoretical results only ensure inversion of the matrix when it the agents are located in a neighbourhood of a Centroidal Voronoi Configuration. In a general situation the matrix can be ill-conditioned, presenting small eigenvalues. This leads to control laws u of TVD-C that have a big norm which exhibit erratic behaviour that can force the agents to escape the domain. This problem can also be found when implementing TVD-D_k for $k \geq 2$, TVD-SP_ϵ and TVD-SSP_ϵ.

This problems are not very frequent, happening in very few iterations, but when they happen they can disrupt the execution of the different algorithms. In Figure 2.4 we can see how $||p-c||^2$, which should decrease exponentially according to (2), is affected by this phenomena. A way to mitigate the erratic behaviour is to reduce the time step, but reducing the time step too much will result in excessive simulation times and real robotic systems have physical limitations on the frequency of operation. This considerations motivate us to introduce a heuristic to trim the components of the control u and avoid having excessively big control laws.

$$u_i' = \begin{cases} u_i & \text{if } |u_i| \leq u_{max,i} \\ u_{max,i}\frac{u_i}{|u_i|} & otherwise \end{cases} \tag{2.24}$$

Here $u_{max,i}$ could be chosen independently for each agent, forbidding agents that are close to the border of the domain to move outside of the domain. Nonetheless in our experiments we will use the same constant u_{max} for all the agents. We can see the effects of the trimming in $||p-c||$ in Figure 2.4 mitigate the erratic behaviour.

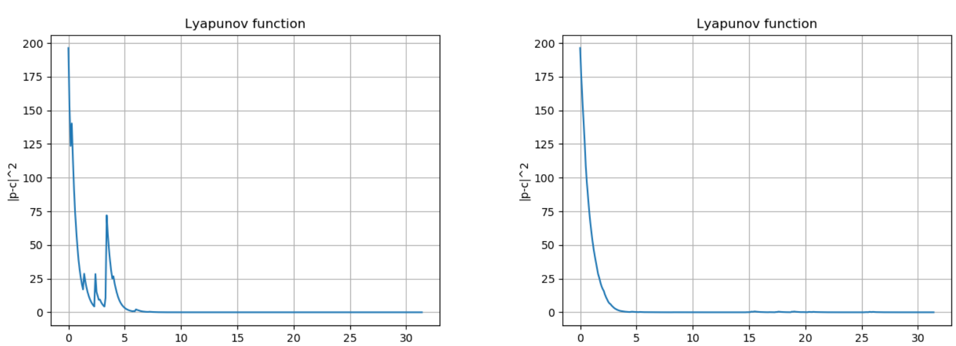

Figure 2.4: Evolution of $||p-c||^2$ with 30 agents in TVD-C without trimming (left) and TVD-C with trimming (right).

We perform experiments with 50 agents imposing $u_{max} = 3$ and show the results on Table 2.2. We use a $\Delta t = 0.1$ and $\Delta\eta = 0.001$. In the results of the table we can see the same patterns we observed with 10 agents but we require smaller values of ϵ to achieve the same results. TVD-SP_ϵ can achieve costs as good as TVD-C by choosing $\epsilon = 0.001$, a perfomance similar to TVD-D_3 can be achieved with $\epsilon = 0.01$. Additionally when we choose a reasonably big value like $\epsilon = 0.1$ the algorithm converges without problems but it achieves a performance comparable to Lloyd algorithm. The other algorithm that we have proposed TVD-SSP_ϵ performs worse than TVD-SP_ϵ, requiring ϵ values 10 times lower than TVD-SP_ϵ to obtain to same performance, as we can see by comparing TVD-SP$_{0.01}$ and TVD-SSP$_{0.001}$.

Table 2.2: Comparison of all the algorithms with 50 agents

Algorithm	ρ_1	ρ_2	ρ_3
Lloyd	33.60	30.40	40.00
TVD-D$_0$	32.67	30.06	37.11
TVD-D$_1$	28.26	26.95	31.84
TVD-D$_2$	26.23	25.56	29.09
TVD-D$_3$	25.07	24.65	27.15
TVD-SP$_{0.1}$	34.09	32.97	41.63
TVD-SP$_{0.01}$	24.67	26.82	30.26
TVD-SP$_{0.001}$	22.91	23.23	22.47
TVD-SSP$_{0.01}$	32.53	31.17	33.69
TVD-SSP$_{0.001}$	27.23	25.59	27.07
TVD-C	22.44	23.32	22.43

Finally, in Figure 2.5 we show how the 50 agents evolve following $\rho_1(x,t)$ using TVD-SP$_{0.001}$, converging to the Centroidal Voronoi configuration and decreasing the coverage cost.

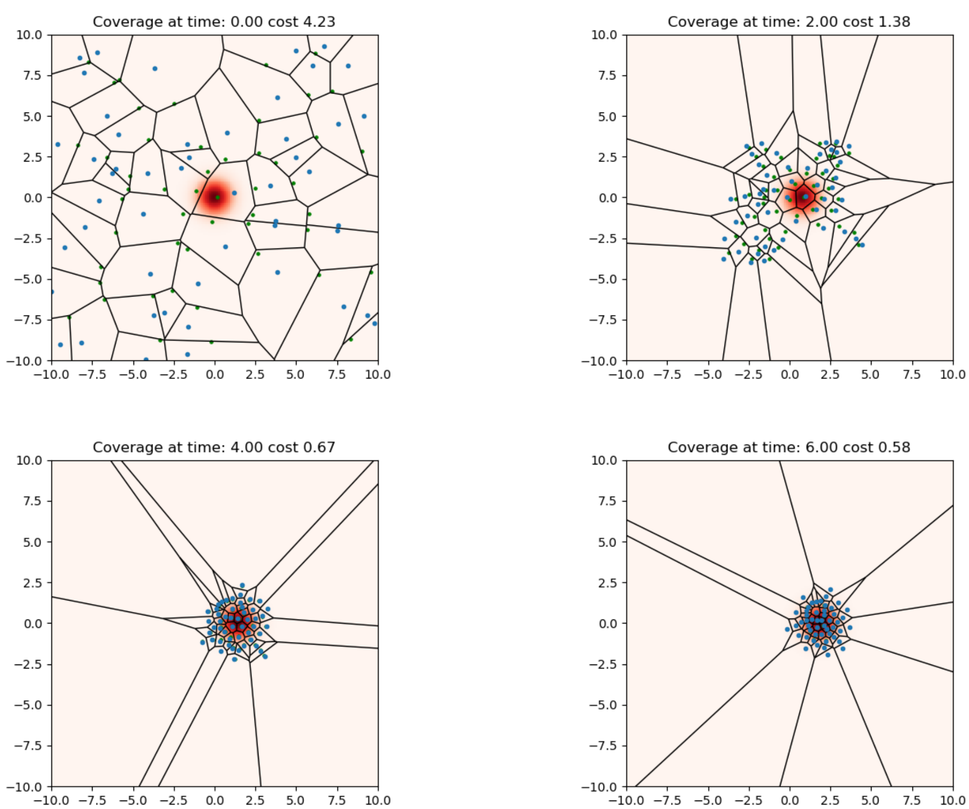

Figure 2.5: Evolution of TVD-SP$_{0.001}$ with 50 agents and target density $\rho_1(x,t)$

Chapter 3

Optimal Transport Theory

Our next objective is to study the transport of Multi-agent robotic systems. To do it we will first introduce the theory of Optimal Transport, which will give us a good framework to study the problem. In this chapter we will give an introductory view on the theory of Optimal Transport. We will follow [49], [50] and [51] and we will present the main results of the Theory of Optimal Transport. We will not prove every proposition or theorem that we state because doing so would require long and technical proofs. We will prove some of the easy results, provide intuition of some of the results and provide references for the more advanced results. Most of the results that we will present in this chapter will be used later to study multi-agent transport.

For the interested reader we recommend [50] and [53] for a complete and detailed approach to Optimal Transport. We also recommend [45] for a computational approach to Optimal Transport. We recommend [51] as an introduction to gradient flows in probability space and [3] for a more in-depth exposition.

3.1 Motivation

Optimal Transport is a theory that studies probability measures, the distance between them and how to efficiently deform a probability measure μ into a target probability measure ν.

The study of the problem started with Gaspard Monge in 1781, when he considered the allocation problem of sending the resources that N different mines in different places had produced to M factories that consume the resources. Monge wanted to find the optimal allocation plan that allowed each mine to send their products to a factory and minimize the transportation cost of the materials.

The problem can be formalized by considering probability measures, and defining a transport cost between probability measures. This cost can be shown to give a metric to the probability space, converting it into a metric space. This allows us to consider functionals on the probability space and minimize functionals in the probability space. This has allowed to study many problems that appear in diverse fields of science and engineering.

3.2 Kantorovich and Monge formulations

We will start introducing the original Monge Formulation and its relaxation into the Kantorovich formulation. The Monge formulation corresponds to the formalization of the allocation problem we have discussed in the motivation section.
Given a space $\Omega \subset \mathbb{R}^N$ we consider two probability measures μ, ν, with support $X = spt(\mu) \subset \Omega$ and $Y = spt(\nu) \subset \Omega$, where the support of a probability measure $spt(\mu)$ is defined as the set of points where the measure is different than 0. We also consider a cost function $c : \Omega \times \Omega \to \mathbb{R}$ and we want to find a map $T : X \to Y$ that transports the mass of μ into ν minimizing the cost of transporting the mass. This can be written as

Definition 3 (Monge formulation)**.**

$$C_M = \inf_{\substack{T:X\to Y \\ T_{\#\mu}=\nu}} \int_X c(x, T(x)) d\mu(x) \tag{3.1}$$

Where the constrain $T_{\#\mu} = \nu$ represents the conservation of mass through the map T, transforming μ into ν. We call $T_{\#\mu}$ the *push-forward* measure of μ through the map T, the measure obtained when we apply the map $X \to Y$ to all the points in the support of μ, and can be defined formally as:

$$T_{\#\mu}(A) = \mu(T^{-1}(A)) \text{ for every measurable set A} \tag{3.2}$$

Now we consider an example that will show us that the Monge formulation quite restrictive.

Example 1. *Consider $\mu = \delta_{x_i}(x)$ a delta function and $\nu(y)$ a probability measure. Then there will only exist a transport map $T : X \to Y$ if $\nu(y) = \delta_{y_i}(y)$ since T is univalued and x_i can only have one image y_i.*

This example shows us that there may be cases where the Monge formulation has no solution. To solve this problem Kantorovich introduced an alternative formulation that generalizes the Monge formulation, allowing the masses to be splitted. In this framework we consider the set of joint probabilities $\Pi(\mu, \nu) \in \mathcal{P}(X \times Y)$ that have as marginals μ and ν and search for the Optimal transport plan $\pi^*(x, y) \in \Pi(\mu, \nu)$ that minimize the cost of transport.

Definition 4 (Kantorovich formulation)**.**

$$C_K = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X\times Y} c(x, y) d\pi(x, y) \tag{3.3}$$

This problem is better posed. We now present some results that relate both formulations

Proposition 3. *Any transport map $T : X \to Y$ induces a transport plan $\pi_T(x, y)$ defined by*

$$\pi_T(x, y) = (id, T)_{\#\mu}$$

Conversely, a transport plan $\pi(x, y)$ induces a transport map T if $\pi(x, y)$ is concentrated on a π-measurable graph Γ, with

$$\Gamma = \{(x, T(x)) : x \in spt(\mu)\}$$

We can also guarantee the existence of transport maps T.

Proposition 4 ([50]). *If μ, ν are two probability measures on $\mathbb{R}^N$ and μ is atomless, then there exist at least a transport map T such that $T_{\#\mu} = \nu$.*

Moreover, it can be proved that the transport plans generated by transport maps T are dense.

Proposition 5 ([50]). *On a set $\Omega \subset \mathbb{R}^N$, the set of plans π_T induced by a transport is dense in the set of plans $\Pi(\mu, \nu)$ whenever μ is atomless.*

Now we introduce a proposition that guarantees the existence of a minimizer $\pi^*(x, y)$.

Proposition 6 ([1]). *Let X and Y be compact metric spaces $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to \mathbb{R}$ continuous. Then the Kantorovich problem admits a solution. Moreover, if the measure μ is atomless (i.e $\mu(\{x\}) = 0$, $\forall x \in X$), we have*

$$\min(KP) = \inf_{T_{\#\mu}=\nu} \int_X c(x, T(x)) d\mu(x)$$

We note that this does not imply the existence of an optimal transport map T^* that minimizes the Monge cost. We provide an example to show this.

Example 2 (3 lines). *We consider $\mu = \mathcal{H}^1|_A$ and $\nu = \frac{1}{2}\mathcal{H}^1|_B + \frac{1}{2}\mathcal{H}^1|_C$. With A, B, C vertical parallel lines, with $x = 0, 1, -1$ and $y \in [0, 1]$ and $\mathcal{H}^1$ is the one dimensional Haussdorf measure. We can divide A into $2n$ segments and assign half of them to B and half to C as we can see in Figure 3.1. If we take the limit $n \to \infty$ we obtain optimal transport cost of 1, corresponding to the horizontal distance between A and B, A and C, but we cannot have a transport plan T attaining this cost, since it would require splitting the mass of each point in A to the right for B and left for C.*

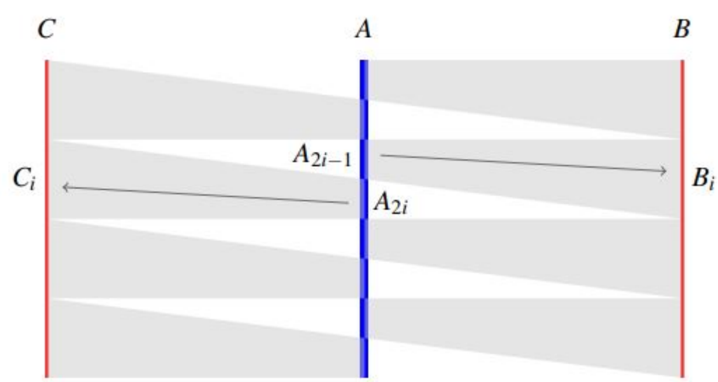

Figure 3.1: Non-existence of transport map example, from [50]

Another important topic is the uniqueness of solutions of the Kantorovich problem. We show now in an example that for a general cost $c : X \times Y \to \mathbb{R}$ there may not be a unique minimizer. We will later see that if we assume some convexity on the cost function it is possible to prove the uniqueness of minimizers.

Example 3 (Book shifter). *We consider the cost* $c(x,y) = |x-y|$ *and the measures* $\mu(x) = \frac{1}{n}\delta_0(x) + \frac{1}{n}\delta_1(x) + \cdots + \frac{1}{n}\delta_n(x)$ *and* $\nu(x) = \frac{1}{n}\delta_1(x) + \frac{1}{n}\delta_2(x) + \cdots + \frac{1}{n}\delta_{n+1}(x)$*, we can think of the problem as wanting to move a set of n books one unit to the right in the real line. We consider two solutions with the same cost:*

1. *Move each book one unit to the right*
2. *Move the leftmost book* $\frac{1}{n}\delta_0(x)$ *to the rightmost place* $\frac{1}{n}\delta_{n+1}(x)$

Both solutions have the same cost of 1, and it is the optimal cost, there is no uniqueness of solution.

3.3 The Wasserstein distance

We will now study the Wasserstein distance, a distance that can be defined in the probability space thanks to Optimal Transport. We will focus on the distances generated by $c = |x-y|^p$ but more general cost functions can be used. We start defining the p-Wasserstein distance W_p as

$$W_p = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} |x-y|^p d\pi(x,y) \right)^{1/p} \tag{3.4}$$

To gain intuition about the Wasserstein distance in Figure 3.3 we show how the Wasserstein distance accounts for the difference between two probability distributions in a different way than traditional L^p distances. The key aspect is that the Wasserstein distance measures the "horizontal" distance by measuring the displacement between x and $T(x)$, while L^p distances measure the "vertical" displacement between the values of two functions in a given point x.

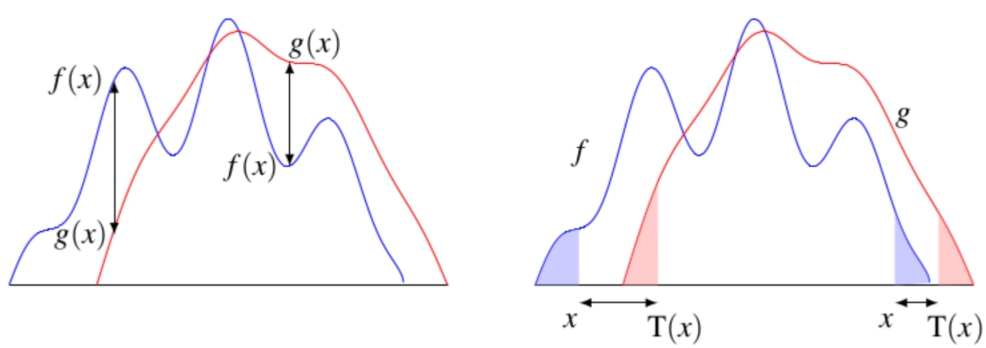

Figure 3.2: Visual explanation of Wasserstein distance by [50]

One of the main benefits of the Wasserstein distance is that it allows to calculate the distance between discrete and continuous probability measures. We introduce an interesting example that shows us how to calculate a Wasserstein distance and also provides an interesting result relating the Wasserstein distance to a target delta function with the p-moments of the source distribution.

Example 4. *We want to calculate the Wasserstein distance $W_p^p(\mu,\nu)$ between $\mu(x)$ an arbitrary probability measure and $\nu(y)=\delta_{x_i}(y)$ a delta function.*
The optimal transport plan between μ and ν is a function that concentrates all the probability of μ into the single point y_i of the support of ν. The transport plan is given by $\pi(x,y)=\mu(x)\times\delta_{y_i}(y)$ and we can calculate the Wasserstein distance as

$$W_p^p(\mu,\nu)=\int_{\mathbb{R}\times\mathbb{R}}|x-y|^p d\pi(x,y)=$$

$$\int_{\mathbb{R}\times\mathbb{R}}|x-y|^p d\mu(x)\delta_{x_i}(y)dy=\int_{\mathbb{R}}|x-x_i|^p d\mu(x)=\mathbb{E}_\mu\left[|x-x_i|^p\right]$$

And $W_p^p(\mu,\nu)=\mathbb{E}_\mu[|x-x_i|^p]$, the p-moment of μ around x_i. This example is particularly interesting if we set $x_i=\bar{\mu}$, the mean value of μ. Then,

$$W_p^p(\mu,\nu)=\mathbb{E}_\mu[|x-\bar{\mu}|^p]$$

And in particular setting $p=2$ we have,

$$W_2^2(\mu,\nu)=\mathbb{E}_\mu[|x-\bar{\mu}|^2]=\mathbb{V}ar_\mu[x]$$

We now prove that the Wasserstein distance establishes a metric in the probability space $\mathcal{P}(\Omega)$.

Proposition 7. *Given a region of space Ω and a metric $c:\Omega\times\Omega\to\mathbb{R}$ then the Wasserstein distance (3.4) defines a metric in $\mathcal{P}(\Omega)$.*

Proof. We start by analyzing the Wasserstein distance

$$W_p=\left(\inf_{\pi\in\Pi(\mu,\nu)}\int_{\Omega\times\Omega}c(x,y)^p d\pi(x,y)\right)^{1/p}$$

From the expression it is easy to see that the distance is non-negative and symmetric. To see that it is non-degenerate we note that since $c(x,y)$ is non-negative it must be 0 in support of the minimizer π^* and $c(x,y)=0$ implies $x=y$.
We will proof the triangular inequality by assuming the existence of and optimal transport maps T between three absolutely continuous measures μ,ν,ρ. A proof that doesn't require this assumption can be found Chapter 5 of [50].
The Wasserstein distance can be written as

$$W_p(\mu,\nu)=\left(\inf_{T_{\#\mu}=\nu}\int_\Omega c(x,T(x))^p d\mu(x)\right)^{1/p}$$

We consider the optimal transport map T^* between μ and ν. Now we consider a third probability measure ρ, and T_1 the optimal transport map from μ to ρ, T_2 the optimal transport map from ρ to ν. Then $T^* = T_2 \circ T_1$ and

$$W_p(\mu,\nu) = \left(\int_\Omega c(x, T_2 \circ T_1(x))^p d\mu(x)\right)^{1/p} \leq \left(\int_\Omega (c(x, T_1(x)) + c(T_1(x), T_2 \circ T_1(x)))^p d\mu(x)\right)^{1/p} \leq$$

$$\leq \left(\int_\Omega c(x, T_1(x))^p d\mu(x)\right)^{1/p} + \left(\int_\Omega c(T_1(x), T_2 \circ T_1(x))^p d\mu(x)\right)^{1/p} = W_p(\mu,\rho) + W_p(\rho,\nu)$$

□

In next proposition we characterize the convergence with respect to the Wasserstein distance. We will not prove it, a proof can be found in [50].

Proposition 8. *Let Ω be compact and $\{\mu_n\}$ be a sequence of measure and μ a candidate limit. Then $W_p(\mu_n, \mu) \to 0$ if and only if $\mu_n \to \mu$ according to the weak convergence, i.e.*

$$\lim_{n\to\infty} \int_\Omega f d\mu_n = \int_\Omega f \mu$$

For all bounded and continuous functions f.

We also remark that the probability space $\mathcal{P}(\Omega)$ is compact with respect to the weak convergence [9], and thanks to last proposition it is also compact with respect to the Wasserstein distance. Then the probability space $\mathcal{P}(\Omega)$ with the Wasserstein distance is a complete metric space, that we will denote as Wasserstein space $\mathbb{W}_p(\Omega)$.
Next we will show some results that characterize the curves in the Wasserstein space. We will not provide proofs of the results that we state, the proofs can be found in [3] and [50].
We start by introducing some notions about curves in metric spaces. We will say that a continuous function $\omega : [t_1, t_2] \to X$ is a curve when it maps an interval $[t_1, t_2] \subset \mathbb{R}$ to a metric space (X, d). We note that the speed $\dot{\omega}(t)$ of the curve has no meaning unless (X, d) is a vector space (the direction is not well-defined), nevertheless we can study the modulus $|\dot{\omega}(t)|$.

Definition 5. *Let (X, d) be a metric space and $\omega : [0, 1] \to X$ a curve valued in a metric space (X, d). We define the metric derivative $|\dot{\omega}(t)|$ of $\omega(t)$ at time t as*

$$|\dot{\omega}(t)| := \lim_{h\to 0} \frac{d(\omega(t+h), \omega(t))}{|h|}$$

provided the limit exists

We now state a theorem on the existence of the metric derivative, which is a generalization of Rademacher Theorem (5).

Theorem 3. *Suppose that $\omega : [0, 1] \to X$ is Lipschitz continuous. Then, the metric derivative $|\dot{\omega}|(t)$ exists for a.e. $t \in [0, 1]$. Moreover we have, for $t < s$,*

$$d(\omega(t), \omega(s)) \leq \int_t^s |\dot{\omega}|(\tau) d\tau$$

We now consider a more general set of curves that don't need to satisfy the Lipschitz continuity

Definition 6. *We will say a curve* $\omega : [0,1] \to X$ *is absolutely continuous whenever there exist a Lebesgue integrable derivative* $g \in L^1([0,1])$ *such that* $d(\omega(t_0), \omega(t_1)) \leq \int_{t_0}^{t_1} g(s)ds$ *for every* $t_0 < t_1$.

Absolutely continuous curves can be reparametrized in time and become Lipschitz continuous [50]. It can then be proved [50], [3], that the absolutely continuous curves of the Wasserstein space $(\mu_t)_{t\in[0,1]}$ are solutions of the continuity equation

$$\delta_t \mu_t + \nabla \cdot (\mu_t \boldsymbol{v_t}) = 0 \tag{3.5}$$

For some vector field $\boldsymbol{v_t}$, which will satisfy $\boldsymbol{v_t} = |\dot{\mu}|(t)$, where $|\dot{\mu}|(t)$ is the metric derivative of the curve $(\mu_t)_{t\in[0,1]}$ with respect to the Wasserstein distance.
Additionally, we can study how a bunch of particles evolve when following the dynamics given in the continuity equation. We will consider some particles initially distributed as $\boldsymbol{x} \sim \mu_0$, then, they will evolve following

$$\begin{aligned} \dot{\boldsymbol{y}}(t) &= \boldsymbol{v_t}(\boldsymbol{y}(t)) \\ \boldsymbol{y}(0) &= \boldsymbol{x} \end{aligned} \tag{3.6}$$

The results that we have presented in this section are very important, allowing us to relate curves in the Wasserstein space with the continuity equation. In our work we will use this relation to develop a multi-agent transport algorithm in Chapter 5.

3.4 Kantorovich dual formulation

In the following section we will take advantage of the linearity of the Kantorovich formulation to study the dual problem of the formulation. The dual formulation of the Kantorovich problem is more simple than the primal formulation and it allows us to define the Kantorovich potentials, which will give us the solution of the problem. Our results are taken mainly from [50]. We start with the Kantorovich formulation (4)

$$C_K = \inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y)d\pi(x,y)$$

Using duality theory we can rewrite the constrain $\pi \in \Pi(\mu, \nu)$ as

$$\sup_{\phi,\psi} \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) - \int_{X\times Y} (\phi(x) + \psi(y))d\pi(x,y) = \begin{cases} 0 \text{ if } \pi \in \Pi(\mu,\nu), \\ +\infty \text{ otherwise} \end{cases}$$

Then we can rewrite the Kantorovich problem as

$$\inf_\pi \int_{X\times Y} c(x,y)d\pi(x,y) + \sup_{\phi,\psi} \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) - \int_{X\times Y} (\phi(x) + \psi(y))d\pi(x,y)$$

Next we will interchange the inf with the sup.[1]

$$\sup_{\phi,\psi} \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) + \inf_{\pi} \int_{X\times Y} (c(x,y) - (\phi(x) + \psi(y)))d\pi(x,y)$$

We rewrite the last term as a constrain using

$$\inf_{\pi} \int_{X\times Y} (c(x,y) - (\phi(x) + \psi(y)))d\pi(x,y) = \begin{cases} 0 \text{ if } \phi(x) + \psi(y) \leq c(x,y) \text{ on } X \times Y \\ -\infty \text{ otherwise} \end{cases} \tag{3.7}$$

This equality comes from the fact that if we have a point (x,y) such that $\phi(x)+\psi(y) > c(x,y)$ then we can concentrate the probability density $\pi(x,y)$ on the point and we get a value of $-\infty$.
From this we get the dual problem.

$$\begin{aligned} C_{DP} = \sup_{\phi,\psi} \quad & \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) \\ \text{s.t.} \quad & \phi(x) + \psi(y) \leq c(x,y) \end{aligned} \tag{3.8}$$

The next big question that comes is whether there exists a solution (ϕ,ψ) of the dual problem. To answer this question we first define the c-transform of a function.

Definition 7. *Given a function $\chi : X \to \bar{\mathbb{R}}$ we define its c-transform or c-conjugate $\chi^c : Y \to \bar{\mathbb{R}}$ by*

$$\chi^c(y) = \inf_{x\in X} c(x,y) - \chi(x)$$

Analogously we define the $\bar{c}-transform$ of $\zeta : Y \to \bar{\mathbb{R}}$ by

$$\zeta^c(x) = \inf_{y\in Y} c(x,y) - \zeta(y)$$

We say that a function $\psi : Y \to \bar{\mathbb{R}}$ is $\bar{c}-$concave if there exist χ such that $\psi = \chi^c$, a function $\phi : X \to \bar{\mathbb{R}}$ is $c-$ concave if there exist ζ such that $\phi = \zeta^c$.

The c-transform is useful because it can be shown that given a pair (ϕ,ψ) if we apply the c-transform we get (ϕ,ϕ^c) with a higher cost and satisfying the same constraints, we could iterate this process and grow the cost infinitely but it can be proved that $\phi^{ccc} = \phi^c$. It can also be proved that the set of functions that are c-concave is compact. This gives us the following result.

Proposition 9 ([50])**.** *Suppose X and Y are compact and c is continuous. Then there exist a solution (ϕ,ψ) to the dual problem with the form $\phi \in c - conc(X), \psi \in \bar{c} - conc(Y)$ and $\psi = \phi^c$. Then,*

$$\max(DP) = \max_{\phi\in c-conc(X)} \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y)$$

Using (7) constrain $\phi \in c - conc(X)$ can be rewritten as $\phi + \phi^c \leq c(x,y)$, giving us

[1] This equivalence is not trivial, we refer the curious reader to section 1.6 of [50]

$$\begin{aligned} C_{DP} = \sup_{\phi} \quad & \int_X \phi(x) d\mu(x) + \int_Y \phi^c(y) d\nu(y) \\ \text{s.t.} \quad & \phi(x) + \phi^c(y) \le c(x,y) \end{aligned} \tag{3.9}$$

With this result we have reduced significantly the complexity of the problem, which now only depends on ϕ.

Definition 8 (Kantorovich potential)**.** *We call ϕ such that the dual problem is maximized Kantorivich potentials, which are equal to a c-concave function $\mu - a.e.$*

Finally we introduce a strong duality result, which will establish the equivalence between the primal and dual formulations of the Kantorovich problem.

Theorem 4 ([50])**.** *Suppose $X, Y \subset \mathbb{R}^N$ and $c : X \times Y \to \mathbb{R}$ is uniformly continuous and bounded. Then the Kantorovich dual problem admits a solution (ϕ, ϕ^c) and we have* $\max(DP) = \min(KP)$

We will now introduce some results that only hold when we add some assumptions to the cost function.

3.4.1 Strictly convex cost functions

In Example 3 we have seen how the optimal transport plan π^* may not be unique and in Example 2 we have seen that there may be situations in which an optimal transport map T^* may not exist. We will now study this in more depth, showing that when $c(x, y) = h(x - y)$ this results hold.
We start with the following proposition, relating the Kantorovich potential with the transport cost c.

Proposition 10. *If the cost c is C^1, ϕ is a Kantorovich potential for the cost c in the transport from μ to ν, and (x_0, y_0) belongs to the support of an optimal transport plan γ, then $\nabla\phi(x_0) = \nabla_x c(x_0, y_0)$, provided ϕ is differentiable at x_0. In particular, the gradients of two different Kantorovich potentials coincide on every point $x_0 \in spt(\mu)$ where both the potentials are differentiable.*

Proof. Using the strong duality of Theorem 4 we can establish that for the optimal Kantorovich potential ϕ^* and the optimal coupling π we have $\max(DP) = \min(KP)$ leading to:

$$\int_X \phi(x) d\mu(x) + \int_Y \phi^c(y) d\nu(y) = \int_{X \times Y} (\phi(x) + \phi^c(y)) d\pi(x,y) = \int_{X \times Y} c(x,y) d\pi(x,y)$$

Then,

$$\phi(x) + \phi^c(y) = c(x,y) \text{ if } (x,y) \in spt(\pi)$$

We also have that ϕ must be dual feasible

$$\phi(x) + \phi^c(y) \le c(x,y) \text{ if } (x,y) \in X \times Y$$

We now fix $y = y_0$ and consider the function

$$x \to \phi(x) - c(x, y_0) \leq -\phi^c(y_0)$$

We know that if $(x_0, y_0) \in spt(\pi)$ the inequality becomes an equality, and $\phi(x) - c(x, y_0)$ achieves a maximum. Then,

$$\nabla_X|_{x=x_0}(\phi(x) - c(x, y_0)) = 0$$

And we end up with

$$\nabla\phi(x) = \nabla_X c(x_0, y_0)$$

□

We note that this equality only holds at the points where ϕ is differentiable. It can be proved that ϕ is Lipschitz [50] and if $spt(\mu) = \Omega$, we can use the Rademanchar theorem to guarantee the differentiability of ϕ.

Theorem 5 (Rademacher). *Let $\Omega \subset \mathbb{R}^N$ be open and $f : \Omega \to \mathbb{R}^m$ be Lipschitz continuous. Then f is differentiable at almost every $x \in \Omega$.*

The equality $\nabla\phi = \nabla_x c$ will help us to find a condition under which there exist an optimal transport map T^*.

Definition 9 (Twist condition). *For $\Omega \subset \mathbb{R}^N$ we say $c : \Omega \times \Omega \to \mathbb{R}$ satisfies the Twist condition when c is differentiable with respect to x at every point and the map $y \to \nabla_x c(x_0, y))$ is injective for every x_0.*

When this condition holds we can deduce that if $(x_0, y_0) \in spt(\pi)$ then y_0 is uniquely defined and there exist an optimal transport map.
Now, if we consider $c(x, y) = h(x, y)$ with h strictly convex, we have $\nabla\phi = \nabla h(x_0 - y_0)$ and since h is strictly convex we can invert ∇h giving us

$$x_0 - y_0 = (\nabla h)^{-1}(\nabla\phi(x_0))$$

This gives rise to the following theorem.

Theorem 6 ([50]). *Given μ and ν probability measures on a compact domain $\Omega \subset \mathbb{R}^n$, with μ is absolutely continuous (with respect to the Lebesgue Measure), and cost given by $c(x, y) = h(x - y)$ with h strictly convex then there exists an optimal transport plan π. It is unique and of the form $(id, T)_{\#\mu}$, provided μ is absolutely continuous with respect to the Lebesgue Measure in Ω and $\partial\Omega$ is negligible. Moreover, there exists a Kantorovich potential ϕ, and T and the potentials ϕ are linked by*

$$T(x) = x - (\nabla h)^{-1}(\nabla\phi(x)).$$

3.4.2 Distances as cost functions

Another interesting case appears when the cost function $c : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ defines a distance over Ω. In this situation we can simplify the dual formulation. The key point is the following lemma.

Lemma 1. *Let c be a metric in Ω and ϕ a Kantorovich potential. Then $\phi^c = -\phi$ and $|\phi(x) - \phi(y)| \leq c(x, y)$ for all $x, y \in \Omega$*

Proof. We consider a conjugate pair $(\phi(x), \psi(y))$ that solves the Kantorovich problem. From Definition 7 of the $c - transform$ we have

$$\psi(y) = \inf_{x \in X} c(x, y) - \phi(x), \quad \phi(x) = \inf_{y \in Y} c(x, y) - \psi(y)$$

And substituting $\psi(y)$ in $\phi(x)$ we obtain

$$\phi(x) = \inf_{y \in Y} c(x, y) - (\inf_{z \in X} c(z, y) - \phi(z)) =$$

$$= \inf_{y \in Y} \sup_{z \in X} c(x, y) - c(z, y) + \phi(z)) \geq \inf_{y \in Y} (c(x, y) - c(z, y)) + \phi(z)) \geq -c(x, z) + \phi(z)$$

Where the last inequality comes from the triangular inequality of the metric c, $c(x, y) - c(y, z) \geq c(x, y) - c(x, y) - c(x, z) = -c(x, z)$. Using $\phi(x) \geq -c(x, y) + \phi(y)$ and the fact that we can swap the points x, y in the expression we obtain

$$|\phi(x) - \phi(y)| \leq c(x, y)$$

Out of this expression we can get

$$-\phi(x) \leq c(x, y) - \phi(y) \leq \phi^c(x)$$

Where we have used that $c(x, y)$ is positive since it is a metric. To find the remaining inequality we take the definition of ϕ^c.

$$\phi^c(x) = \inf_{y} c(x, y) - \phi(y) \leq c(x, y) - \phi(y) \leq -\phi(y)$$

And taking $y = x$ we get $\phi^c(x) \leq -\phi(x)$ □

Assuming that the cost is a metric and using Lemma 1 we can write a simplified version of the Kantorovich formulation.

$$\begin{aligned} C_{DP} = \sup_{\phi} \quad & \int_{\Omega} \phi(x) d\mu(x) - \int_{\Omega} \phi(y) d\nu(y) \\ \text{s.t.} \quad & |\phi(x) - \phi(y)| \leq c(x, y) \end{aligned} \tag{3.10}$$

Now we will focus our attention on Riemmanian metrics.

Definition 10. *We say a metric $c : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ is Riemannian or conformal to the Euclidean distance if there exist a continuously differentiable conformal factor $\xi(x) > 0$ such that*

$$c(x, y) = \inf_{\substack{\gamma(t):[0,1] \to \Omega \\ \gamma(0)=x, \gamma(1)=y}} \int_0^1 \xi(\gamma(t)) ||\dot{\gamma}(t)|| dt$$

Where the infimum is taken over the curves $\gamma(t)$ with endpoints $\gamma(0) = x$, $\gamma(1) = y$.

With the Riemmanian metric assumption it is easy to check that the Kantorovich potential is Lipschitz in $\mathring{\Omega}$. We consider $x \in \mathring{\Omega}$, $y \in B(x,\epsilon) \subset \mathring{\Omega}$ and a straight line γ connecting x and y.

$$|\phi(x) - \phi(y)| \leq c(x,y) \leq \int_0^1 \xi(\gamma(t))||\dot{\gamma}(t)||dt = \gamma(z)\int_0^1 ||\dot{\gamma}(t)||dt = \gamma(z)||x-y||$$

Where we have used the Mean Value Theorem with $z = mx + (1-m)y$ with $m \in [0,1]$. Since ϕ is Lipschitz from the Rademacher Theorem (5) we know that it is differentiable almost everywhere. With this considerations we will rewirte the constrain.

Proposition 11. *The constrain* $|\phi(x) - \phi(y)| \leq c(x,y)$ *is equivalent to* $||\nabla\phi(x)|| \leq \xi(x)$

Proof. Let $x \in \mathring{\Omega}$ and $y \in B(x,\epsilon) \subset \mathring{\Omega}$. Then, using the Mean Value Theorem, there exist $z_1 = m_1 x + (1-m_1)y$, $z_2 = m_2 x + (1-m_2)y$ with $m_1, m_2 \in [0,1]$ such that

$$\frac{|\nabla\phi(z_1)\cdot(x-y)|}{||x-y||} = \frac{|\phi(x)-\phi(y)|}{||x-y||} \leq \frac{c(x,y)}{||x-y||} \leq \frac{\xi(z_2)||x-y||}{||x-y||}$$

If we take $y = x + tv$ and take $t \to 0$ then $z_1, z_2 \to x$ and we will obtain

$$\frac{\nabla\phi(x)\cdot v}{||v||} \leq \xi(x)$$

For every $v \in \mathcal{T}_x(\Omega)$, the tangent space at point $x \in \mathring{\Omega}$. This gives us the inequality $||\nabla\phi|| \leq \xi(x)$.
To prove the converse we assume $\nabla\phi \leq \xi(x)$ and we consider the geodesic γ of the cost c that connects $\gamma(0) = x$ and $\gamma(1) = y$. Then,

$$|\phi(x)-\phi(y)| = \int_0^1 \nabla\phi\cdot\dot{\gamma}(t)dt \leq \int_0^1 ||\nabla\phi||\,||\dot{\gamma}(t)||dt \leq \int_0^1 \xi(\gamma(t))||\dot{\gamma}(t)||dt = c(x,y)$$

□

This gives rise to the following formulation of Optimal Transport

$$\begin{aligned} C_{DP} = \sup_{\phi} \quad & \int_\Omega \phi(x)d\mu(x) - \int_\Omega \phi(y)d\nu(y) \\ \text{s.t.} \quad & |\nabla\phi(x)| \leq \xi(x) \end{aligned} \tag{3.11}$$

The two formulations that we have presented in this section have been used for multi-agent transport in [30]. We will talk more about this approach in Chapter 5

3.5 Fluid Mechanics interpretation of Optimal Transport

In next section we will provide a fluid mechanics interpretation of Optimal Transport. We will only give a very superficial treatment on the topics exposed, more information can be found in Chapter 4 of [50].
In fluid mechanics there are two equivalent formalisms to describe the motion of particles, the Lagrangian point of view and the Eulerian point of view. In the Lagrangian point of view we label some particles at an initial time and follow the motion (position and velocity) of the particles as time goes by. The Monge and the Kantorovich formulations of optimal transport can be considered Lagrangian, since we start with a measure μ and follow how the different points in the domain evolve. If there is a unique image the evolution is given by a transport map T and if there are more than one images we have a transport plan π that allows us to split the particle and the mass.
In the Eulerian point of view instead of following particles we focus on regions of space, we describe the evolution by providing the velocity, the density of mass and/or directional flow rate in each point of the space at every time. We distinguish two models within the Eulerian point of view, static and dynamic. In the dynamic model we use the density $\rho(x,t)$ and the velocity field $\boldsymbol{v}(x,t)$ to describe the motion. We start with an initial density ρ_0, the particles are initially distributed following $\boldsymbol{x} \sim \rho_0$ and the evolution is given by

$$\begin{aligned} \dot{\boldsymbol{y}}(t) &= \boldsymbol{v_t}(\boldsymbol{y}(t)) \\ \boldsymbol{y}(0) &= \boldsymbol{x} \end{aligned} \tag{3.12}$$

The particles evolve following the map $\boldsymbol{Y_t}(x) = \boldsymbol{y_x}(t)$, the measure evolves with the push-forward $\rho_t = (\boldsymbol{Y_t})_{\#\rho_0}$ and ρ_t, $\boldsymbol{v_t}$ solve the continuity equation

$$\delta_t \rho_t + \nabla \cdot (\rho_t \boldsymbol{v_t}) = 0$$

We have already talked about this formulation when we considered absolutely continuous curves in the Wasserstein space, which also satisfied the continuity equation. More information about the continuity equation can be found in [2].
We will now introduce the static Eulerian framework. In this framework we consider a flow rate $\boldsymbol{w} : \mathcal{P}(\Omega) \to \mathbb{R}^N$, which is a vector measure. $\boldsymbol{w}$ has units of density times velocity. We consider μ as a source of flow and ν as a sink of flow. Then we can impose conservation of flow in every measurable set A.

$$\int_{\delta A} \boldsymbol{w} \cdot \boldsymbol{n} dx = \int_A (\mu - \nu) dx$$

The flow that crosses through the border of A is equal to the difference of masses in the set A. Using the Divergence theorem, we have,

$$\int_A \nabla \cdot \boldsymbol{w} dx = \int_A (\mu - \nu) dx$$

This corresponds to a continuity equation,

$$\nabla \cdot \boldsymbol{w} = \mu - \nu$$

In the following subsections we show how this Eulerian transport models can give us Optimal Transport formulations.

3.5.1 Benamou-Brenier formulation

We return to the dynamic formulation given by $(\rho_t, \boldsymbol{v_t})$ and the continuity equation

$$\delta_t \rho_t + \nabla \cdot (\rho_t \boldsymbol{v_t}) = 0$$

This formulation not only describes curves in the Wasserstein space, if the domain $\Omega \subset \mathbb{R}^N$ is convex and compact and $p > 1$ we can Optimize the Transport cost searching for a constant speed geodesic solving the Benamou-Brenier formulation [8].

$$W_p^p(\mu, \nu) = \min_{\rho_t, \boldsymbol{v_t}} \left\{ \int_0^1 \int_\Omega |\boldsymbol{v_t}|^p d\rho_t dt : \delta_t \rho_t + \nabla \cdot (v_t \rho_t) = 0, \rho_0 = \mu, \rho_1 = \nu \right\} \tag{3.13}$$

Solving this problem we can find $(\rho_t, \boldsymbol{v_t})$ belonging to the Optimal Transport geodesic, and from that find the Optimal Transport plan. This has allowed the creation of numerical algorithms to solve optimal transport, out of them we highlight the Computational Fluid dynamics approach of [8]. The ideas described in this section were used by [6] in order to implement a multi-agent transport algorithm.

3.5.2 Beckmann's problem

In this section we will use the static Eularian formulation of transport and we introduce the Beckmann minimal flow problem. This problem is sometimes used in traffic models and can be related to the Kantorovich formulation of Optimal Transport.

$$\begin{aligned} C_{BP} = \inf_{\boldsymbol{w}:\Omega \to \mathbb{R}^d} \quad & \int_\Omega \xi(x) |\boldsymbol{w}(x)| dx \\ \text{s.t.} \quad & \nabla \cdot \boldsymbol{w} = \mu - \nu \end{aligned} \tag{3.14}$$

We will prove the equivalence of the Beckmann problem with the Kantorovich formulation when $c(x, y) = \inf_\gamma \int_0^1 \xi(\gamma(t)) |\dot{\gamma}(t)| dt$, which is a generalization of the cost $c(x, y) = |x - y|$.

We start by considering the constrain $\nabla \cdot w = \mu - \nu$, which taken in the weak sense becomes

$$-\int_\Omega \nabla \phi \cdot d\boldsymbol{w} = \int_\Omega \phi d(\mu - \nu), \, \forall \phi \in C^1(\bar{\Omega})$$

If we take the supremum over ϕ we can rewrite the constrain

$$\sup_{\phi} \int_{\Omega} \nabla\phi \cdot d\boldsymbol{w} + \int_{\Omega} \phi d(\mu - \nu) = \begin{cases} 0 \text{ if } \nabla \cdot \boldsymbol{w} = \mu - \nu \\ +\infty \text{ otherwise} \end{cases}$$

We can rewrite the objective function of the Beckmann problem as

$$\inf_{\boldsymbol{w}} \int_{\Omega} \xi(x)|\boldsymbol{w}(x)|dx + \sup_{\phi} \int_{\Omega} \nabla\phi \cdot d\boldsymbol{w} + \int_{\Omega} \phi d(\mu - \nu)$$

Interchanging the supremum and the infimum we get

$$\sup_{\phi} \int_{\Omega} \phi d(\mu - \nu) + \inf_{\boldsymbol{w}} \int_{\Omega} \xi(x)|\boldsymbol{w}(x)|dx + \int_{\Omega} \nabla\phi \cdot d\boldsymbol{w}$$

The infimum term has the values

$$\inf_{\boldsymbol{w}} \int_{\Omega} \xi(x)|\boldsymbol{w}(x)|dx + \int_{\Omega} \nabla\phi \cdot d\boldsymbol{w} = \begin{cases} 0 \text{ if } |\nabla\phi(x)| \leq \xi(x) \\ -\infty \text{ otherwise} \end{cases}$$

Giving us the dual formulation of the Beckmann problem

$$\begin{aligned} C_{DBP} = \sup_{\phi} \quad & \int_{\Omega} \phi(x)d\mu(x) - \int_{\Omega} \phi(y)d\nu(y) \\ \text{s.t.} \quad & |\nabla\phi(x)| \leq \xi(x) \text{ in } \Omega \end{aligned} \tag{3.15}$$

We now recall the Kantorovich dual formulation when c is a cost (3.10)

$$\begin{aligned} C_{DP} = \sup_{\phi} \quad & \int_{\Omega} \phi(x)d\mu(x) - \int_{\Omega} \phi(y)d\nu(y) \\ \text{s.t.} \quad & |\phi(x) - \phi(y)| \leq c(x,y)\, \forall x, y \in \Omega \end{aligned} \tag{3.16}$$

The objective function of the Dual Beckmann problem and the Dual Kantorovich problem are the same. If we consider a Riemannian metric given by

$$c(x,y) = \inf_{\substack{\gamma(t):[0,1]\to\Omega \\ \gamma(0)=x,\, \gamma(1)=y}} \int_0^1 \xi(\gamma(t))|\dot{\gamma}(t)|dt$$

Then the constrains are also equivalent as we have already shown in (3.11). We note that here we are using the L^1 norm $|\cdot|$. The equivalence between the Beckmann formulation and the Kantorovich formulation only holds when the cost is $c(x,y) = |x-y|$ or a conformal cost with strictly positive conformal factor $\xi(x) > 0$.

During our early research we studied the possibility of using Beckmann's problem to solve a multi-agent transport problem. We were interested in the Beckmann problem because it can be discretized in a graph, giving us the well known Network Flows problem of finding the Minimum cost Flow over a graph. In our approach we tried to follow up on the results of [23] and provide some improvements. After some research we decided to terminate this line of research. We won't present the results of this research.

3.6 Functionals over probabilities and Gradient Flows

We will now talk about how we can define functionals on the probability space and mimimize them. This is one of the main uses of Optimal Transport. Our treatment will not be complete and we will particularly focus on the Wasserstein distance to a target probability density, $W_p^p(\cdot, \nu)$. For a more detailed treatment we recommend reading Chapters 7 and 8 of [50] and [51]. For a more in depth treatment we also recommend [3].
In the probability space $\mathcal{P}(\Omega)$ we can define functionals $\mathcal{F} : \mathcal{P}(\Omega) \to \mathbb{R}$. We start showing different examples of functionals that can be defined.

- Integrals of a function (potential energy)

$$\mathcal{V}(\mu) = \int_\Omega V d\mu$$

- Double integrals of a function in $\mu \times \mu$ (interaction energy)

$$\mathcal{W}(\mu) = \int_\Omega W d\mu \times d\mu$$

- The Wasserstein distance W_p^p to a target probability measure $W_p^p(\mu, \nu)$

$$\mu \to W_p^p(\mu, \nu)$$

- The integral of a function of the density

$$\mathcal{G}(\mu) = \begin{cases} \int_\Omega g(\rho(x))dx & \text{if } \mu = \rho dx \\ +\inf & \text{otherwise} \end{cases}$$

- The sum of a function of the masses

$$\mathcal{H}(\mu) = \begin{cases} \sum_i h(a_i) & \text{if } \mu = \sum_i a_i \delta_{x_i} \\ +\inf & \text{otherwise} \end{cases}$$

We recommend reading [48] to see how this functionals can be used to model different problems from different fields.

3.6.1 Continuity of functionals

After having defined functionals over the space of probabilities we are interested in minimizing them. But first we must ask more basic questions. We will now concern ourselves

with the existence of minimums and the convexity of the functionals. To find minimizers one can rely on the fact that $\mathcal{P}(\Omega)$ is compact if Ω is compact. Then if the functional $\mathcal{F}(\mu)$ is continuous from the Extreme Value Theorem we know that it will attain a maximum and a minimum in $\mathcal{P}(\Omega)$.
We note that when we talk about a continuous functional we refer to $\mu_n \to \mu \implies \mathcal{F}(\mu_n) \to \mathcal{F}(\mu)$, where $\mu_n \to \mu$ is the weak convergence of probability measures. We will work with continuous functionals but we remark that this results can be relaxed considering semi-continuous functions.

Definition 11. *We say a function $f : X \to \mathbb{R} \cup \{+\infty, -\infty\}$ is upper semi-continuous if*

$$\limsup_{x \to x_0} f(x) \leq f(x_0)$$

We say a function $f : X \to \mathbb{R} \cup \{+\infty, -\infty\}$ is lower semi-continuous if

$$\liminf_{x \to x_0} f(x) \geq f(x_0)$$

The Extreme Value Theorem will ensure that an upper semi-continuous function achieves a maximum in a compact set and a lower semi-continuous function achieves a minimum in a compact set.
We will now focus on the functional $W_p^p(\mu, \nu) = inf_{\pi \in \Pi(\mu,\nu)} \int_\Omega |x - y|^p d\pi(x, y)$ and the generalization for an arbitrary cost $c : \Omega \times \Omega \to \mathbb{R}$. $\mathcal{T}_c(\mu, \nu) = inf_{\pi \in \Pi(\mu,\nu)} \int_\Omega c(x, y) d\pi(x, y)$.

The first step is to prove the continuity of the functionals.

Proposition 12 ([50])**.** *For any $p \leq +\infty$, the Wasserstein distance $W_p(\cdot, \nu)$ to any fixed measure $\nu \in \mathcal{P}(\Omega)$ is continuous with respect to the weak convergence provided Ω is compact. More generally, for any continuous cost $c : \Omega \times \Omega \to \mathbb{R}$ the functional $\mathcal{T}_c(\cdot, \nu)$ is continuous if Ω is compact.*

3.6.2 First variation

We will continue by studying the derivatives of functionals. To define the notion of derivative of a functional we will use the first variation of a functional $\frac{\delta \mathcal{F}}{\delta \rho}(\rho)$, which corresponds to differentiating the functional with respect to perturbations in the probability measure μ. We provide the definition of first variation from [51].

Definition 12. *Given a functional $\mathcal{F} : \mathcal{P}(\Omega) \to \mathbb{R}$ we call $\frac{\delta \mathcal{F}}{\delta \rho}(\rho)$, if it exists, the unique (up to additive constants) function such that*

$$\frac{d}{d\epsilon} \mathcal{F}(\rho + \epsilon\chi)_{|\epsilon=0} = \int_\Omega \frac{\delta \mathcal{F}}{\delta \rho}(\rho) d\chi$$

for every perturbation χ such that, at least for $\epsilon \in [0, \epsilon_0]$ the measure $\rho + \epsilon\chi$ belongs to $\mathcal{P}(\Omega)$.

Example 5. *We provide two examples of how to compute the first variation*

$$\mathcal{F}(\rho) = \int_\Omega f(\rho(x))dx \implies \frac{\delta \mathcal{F}}{\delta \rho}(\rho) = f'(\rho)$$

$$\mathcal{V}(\rho) = \int_\Omega V(x)d\rho(x) \implies \frac{\delta \mathcal{V}}{\delta \rho}(\rho) = V$$

Now we will focus on the transport cost $\mathcal{T}_c(\mu,\nu)$ and its first variation. If we express the transport cost $\mathcal{T}_c(\mu,\nu)$ in the dual formulation

$$\mathcal{T}_c(\mu,\nu) = \sup_{\phi \in c-conc(\Omega)} \int_\Omega \phi d\mu + \int_\Omega \phi^c d\nu$$

If we forget about the supremum the in the expression the first variation with respect to μ will be given by the function ϕ inside the integral, as we have seen in the example. Then the first variation of $\mathcal{T}_c(\mu,\nu)$ will be given by the Kantorovich potentials which maximize the objective function. This leads to the following proposition.

Proposition 13 ([50])**.** *Let $\Omega \subset \mathbb{R}^N$ be compact and $c : \Omega \times \Omega \to \mathbb{R}$ be continuous. Then the functional $\mu \to \mathcal{T}_c(\mu,\nu)$ is convex, and its sub-differential at μ is equal to the set of Kantorovich potentials $\{\phi \in C^0(\Omega) : \int_\Omega \phi d\mu + \int_\Omega \phi^c d\nu = \mathcal{T}_c(\mu,\nu)\}$. Moreover, if the Kantorovich potential is unique up to constants there exists a first variation and it is given by $\frac{\delta \mathcal{T}_c(\cdot,\nu)}{\delta \rho}(\mu) = \phi$.*

In this proposition we have used the notion of sub-differential, that we have not defined previously. A sub-differential is a generalization of a derivative for convex functions that allows multiple values for the derivative.

Definition 13. *For every convex function f we define its sub-differential at x as the set*

$$\delta f(x) = \{p \in \mathbb{R}^N : f(y) \geq f(x) + p \cdot (y - x) \quad \forall y \in \mathbb{R}^N\}$$

Fortunately for us, the assumption that the Kantorovich potential is unique is not very restrictive, from (10), $\nabla \phi = \nabla_x c(x_0, y_0)$ with $(x_0, y_0) \in spt(\pi^*)$, it is easy to see that the Kantorovich potential will be unique up to additive constants if the cost is C^1.

Proposition 14. *If Ω is the closure of a bounded connected open set, the cost c is C^1 and at least one of the measures μ, ν is supported on the whole Ω, then the c-concave Kantorovich potential from μ to ν is unique up to additive constants.*

Additionally, strict convexity of $\mathcal{T}_c(\mu,\nu)$ may be proved.

Proposition 15 ([50])**.** *If $\nu(x) = \rho(x)dx$ has a density and the cost c satisfies the Twist condition (9), $\mathcal{T}_c(\mu,\nu)$ is strictly convex.*

In our work when we talk about Kantorovich potentials we will always assume that they are unique up to additive constants.

3.6.3 Convexity

We will now focus on the notion of convexity, we will start by defining convexity in $\mathbb{R}^N$.

Definition 14. *Let $\Omega \subset \mathbb{R}^N$ be a convex set and $f : \Omega \to \mathbb{R}$, we say f is convex if for all $x, y \in \Omega$ and $t \in [0, 1]$ we have*

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

We note that in order to define convexity we have used a notion of interpolation, considering the linearly interpolated points $z = (1-t)x + ty$ with $t \in [0, 1]$.
When we talk about convexity in the probability space we must first clarify what notion of interpolation we are using. In last section we have talked about the convexity of the transport cost $\mathcal{T}_c(\mu, \nu)$ using the usual interpolation

$$\mu_t = (1-t)\mu_0 + t\mu_1$$

This interpolation can be inconvenient in the context of Optimal Transport, if we consider $\mu_0 = \delta_{x_0}$ and $\mu_1 = \delta_{x_1}$ then

$$\mu_t = (1-t)\delta_{x_0} + t\delta_{x_1}$$

A more suitable interpolation will be given by

$$\mu_t = \delta_{(1-t)x_0 + tx_1}$$

We will call this interpolation the displacement interpolation.

Definition 15. *Let $\mu, \nu \in \mathcal{P}(\Omega)$ such that there exist an optimal transport map $T : \Omega \to \Omega$ from μ to ν that minimizes the Wasserstein-2 distance. Then the displacement interpolation is given by*

$$\mu_t = ((1-t)id + tT)_{\#\mu} \quad \forall t \in [0, 1]$$

We remember from (6) that if μ is atomless there will exist an optimal transport map T.

We note that this interpolation interpolates the points following the geodesics of the metric space. We will now define geodesic convexity in an arbitrary geodesic metric space (a metric space where the distance comes from the length of the geodesics).

Definition 16. *In a geodesic metric space X, we define $F : X \to \mathbb{R} \cup \{+\infty\}$ to be geodesically convex if for every two points $x_0, x_1 \in X$ there exists a constant speed geodesic γ connecting $\gamma(0) = x_0$ to $\gamma(1) = x_1$ such that $[0, 1] \ni t \to F(\gamma(t))$ is convex.*

We now want to prove that the Wasserstein distance $W_2^2(\mu, \nu)$ is geodesically convex using the notion of displacement interpolation. Unfortunately for us, there exists counter examples to the geodesic convexity of $W_2^2(\mu, \nu)$ in $\mathbb{W}_2$ [50].
This is quite counter-intuitive and raises some problems in the study of the Wasserstein space and gradient flows. To bypass this problems we can use generalized geodesics, which we will define next.

Definition 17. *If we fix $\rho \in \mathcal{P}(\Omega)$, for every point $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ we call generalized geodesic between μ_0 and μ_1 with base ρ in $\mathbb{W}_2(\Omega)$ the curve $\mu_t = ((1-t)T_0 + tT_1)_{\#\rho}$, where T_0 is the optimal transport map for the cost $c(x,y) = ||x-y||^2$ from ρ to μ_0 and T_1 from ρ to μ_1.*

Then, $t \to W_2^2(\mu_t, \rho)$ satisfies

$$W_2^2(\mu_t, \rho) \leq \int |(1-t)T_0(x) + tT_1(x)|^2 d\rho(x) \leq$$

$$\leq (1-t)\int |T_0(x) - x|^2 d\rho(x) + t\int |T_1(x) - x|^2 d\rho(x) = (1-t)W_2^2(\mu_0, \rho) + tW_2^2(\mu_1, \rho)$$

3.6.4 Gradient flows in the Euclidean space

In the next sections we will study how to minimize functionals. We start by defining gradient flows, which are the continuous time systems corresponding to the gradient descend algorithm. We start by describing gradient flows in $\mathbb{R}^N$, which we will later extend to $\mathcal{P}(\Omega)$.

Definition 18. *Let $F : \mathbb{R}^N \to \mathbb{R}$ and a point $x_0 \in \mathbb{R}^N$. We consider the dynamic system*

$$\begin{aligned} \dot{x}(t) &= -\nabla F(x(t)) \\ x(0) &= x_0 \end{aligned} \tag{3.17}$$

It can be easily proved that this dynamic system minimizes the function F. For simplicity we show it when F is differentiable

$$\frac{dF}{dt}(x(t)) = \frac{dF}{dx} \cdot \dot{x}(t) = -\|\nabla F(x(t))\|^2 \leq 0$$

And we will converge to $\{x \in \mathbb{R}^N : \nabla F(x) = 0\}$. Next we will discretize the gradient flow in time, to do it we consider the proximal gradient descend recursion given by

$$x_{k+1}^\tau = \arg\min_x F(x) + \frac{|x - x_k^\tau|^2}{2\tau} \tag{3.18}$$

This recursion is well defined when F is convex, but can be relaxed to λ-convex functions [50], but we will work with convex functions. It is also important to note that the proximal gradient recursion (3.18) can be interpreted as an implicit Euler integration.

$$\arg\min_x F(x) + \frac{|x - x_k|^2}{2\tau} \implies \nabla F(x_{k+1}) + \frac{x_{k+1} - x_k}{\tau} = 0 \implies$$
$$x_{k+1} = x_k - \tau \nabla F(x_{k+1})$$

3.6.5 Gradient flows in $\mathbb{W}_2$

With the previous section results we can study the minimization functionals with gradient flows in the Wasserstein space $\mathbb{W}_2$. We consider a functional $\mathcal{F} : \mathcal{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$, which we will assume continuous with respect to the weak convergence of probabilities. We start by solving the proximal recursion

$$\rho_{k+1}^{\tau} = \arg\min_{\rho} \mathcal{F}(\rho) + \frac{W_2^2(\rho, \rho_k^{\tau})}{2\tau}$$

When we achieve the minimimum the first variation or derivative is equal to 0, this gives us

$$\frac{\delta \mathcal{F}}{\delta \rho}(\rho) + \frac{\phi_{\rho \to \rho_k}}{\tau} = constant$$

Where we have used (13) to write the first variation of the Wasserstein distance as the Kantorovich potential $\phi_{\rho \to \rho_k}$. Now we recall from (6) that $T(x) = x - \nabla \phi(x)_{\rho \to \rho_k}$, giving us

$$\frac{T(x) - x}{\tau} = -\frac{\nabla \phi_{\rho \to \rho_k}}{\tau} = \nabla \left(\frac{\delta \mathcal{F}}{\delta \rho}(\rho) \right)$$

We can define $\boldsymbol{v} = \frac{x - T(x)}{\tau} = -\nabla \left(\frac{\delta \mathcal{F}}{\delta \rho}(\rho) \right)$. If we take $\tau \to 0$ the sequence $\{\rho_k\}_k$ becomes a continuous curve, which as we have mentioned previously will be defined by the continuity equation

$$\delta_t \rho_t + \nabla \cdot (\rho_t \boldsymbol{v_t}) = 0$$

Where $\boldsymbol{v_t} = -\nabla \left(\frac{\delta \mathcal{F}}{\delta \rho}(\rho) \right)$. In particular when we consider the Wasserstein distance $W_2^2(\mu, \nu)$ to a target probability measure ν as the functional to be minimized we have $\boldsymbol{v_t} = -\nabla \phi_{\mu \to \nu}$, and the transport generated by the gradient potential is a gradient flow if the Wasserstein distance.

Remark 2. *We have developed this section in the space $\mathbb{W}_2$ but the same techniques can be generalized to the probability space $\mathcal{P}(\Omega)$ with a more generalized transport cost $\mathcal{T}_c(\mu, \nu)$. We leave the generalization to other spaces for more advanced texts [3].*

We will use the results of this section in Chapter 5 to develop a multi-agent transport algorithm.

3.6.6 The Fokker-Planck equation in $\mathbb{W}_2$

We will devote this small section to comment a particular gradient flow as an example of the uses of Optimal Transport. The proofs of the results we state can be found in [50].
We start by defining the functional

$$J(\rho) = \int_{\Omega} \rho \log \rho + \int_{\Omega} V d\rho$$

Then, we can define a gradient flow on the functional and it can be proven, using the same techniques that we have already discussed that the gradient flow is given by the Fokker-Plank equation

$$\begin{aligned} \delta_t\rho - \Delta\rho + \nabla \cdot (\rho \nabla V) = 0 \\ \rho(0) = \rho_0 \end{aligned} \tag{3.19}$$

Which corresponds to the evolution of the position of a group of particles initially distributed according to $x_0 \sim \rho_0$ and following the evolution

$$dX_t = -\nabla V(X_t)dt + dB_t$$

Where dB_t accounts for Brownian motion.
This result is very important bacause if we take $V = 0$ the function we are minimizing is the entropy

$$S(\rho) = \int_\Omega \rho \log \rho$$

And the diffusion equation is a gradient flow

$$\delta_t\rho - \Delta\rho = 0$$

And the evolution of a group of particles under Brownian motion $dX_t = dB_t$ generates a gradient flow on the entropy of the system. With this result we conclude our introduction to Optimal Transport.

Chapter 4

Optimal Transport and Coverage Control

In this short chapter we will study the transport problem in the case where one probability measure is discrete and the other has a density. Starting from the Kantorovich dual form we will find the Optimal Transport plan which will have a structure given by Voronoi Power Diagrams.
Then we will talk about the Coverage problem and discuss the relations between Optimal Transport and Coverage Control. We will see that the locational cost used in Coverage Control cost is a relaxation of the Optimal Transport cost but there are some differences between the two problems. We will also see that Coverage Control is not well-posed to be analyzed macroscopically.

4.1 Semi-discrete Optimal Transport

In this section we follow [50] to find the transport plan of the Semi-Discrete Optimal Transport. We consider $\mu(x) = \sum_{i=1}^{n} a_i \delta_{x_i}(x)$ discrete and $\nu(x) = \rho(x)dx$ continuous. We start recalling the definition of the Kantorovich dual formulation in (9).

$$C_{DP} = \sup_{\phi \in c-conc} \quad \sum_{i=1}^{n} a_i \phi(x_i) + \int_{\Omega} \phi^c(y) d\nu(y)$$

We recall from our calculations of the Kantorovich dual form in 3.7 that

$$\inf_{\pi} \int_{\Omega \times \Omega} (c(x,y) - (\phi(x) + \phi^c(y))) d\pi(x,y) = \begin{cases} 0 \text{ if } \phi(x) + \phi^c(y) \le c(x,y) \text{ on } X \times Y \\ -\infty \text{ otherwise} \end{cases}$$

From that we know that $\phi(x) + \phi^c(y) = c(x,y)$ on $spt(\pi)$, which are the pairs (x,y) where we have mass transport and we can find the points $y \in \Omega$ that are mapped to a point x_i as

$$c(x_i, y) - \phi(x_i) = \phi^c(y)$$

Recalling the definition of $c - transform$ from 7 we have

$$c(x_i, y) - \phi(x_i) = \phi^c(y) = \inf_{x_j \in \Omega} c(x_j, y) - \phi(x_j)$$

We can define the set,

$$V_{\phi_i} = \{y \in \Omega : c(x_i, y) - \phi(x_i) \leq c(x_j, y) - \phi(x_j) \quad \forall j\}$$

Which corresponds to the set of points $y \in \Omega$ that satisfy the equality $\phi(x) + \phi^c(y) = c(x, y)$ and the points $y \in V_{\phi_i}$ will be mapped x_i. Additionally we note that the sets V_i correspond to a Voronoi Power Cell 2 with weights $(\phi_1, \dots, \phi_n)$.
Next, we want to find the optimal weights $(\phi_1, \dots, \phi_n)$, which will assigns to the Voronoi Power Cell V_{ϕ_i} the same amount of mass as the discrete value $\mu(x_i) = a_i$. This can be done by solving the finite dimensional problem of maximizing the objective in the Kantorovich dual formulation

$$C_{DP} = \max_{\phi} \sum_{i=1}^{n} a_i \phi(x_i) + \int_{\Omega} \phi^c(y) d\nu(y)$$

From the definition of $\phi^c(y)$ we can deduce it's derivative as

$$\frac{\delta \phi^c}{\delta \phi_j} = \begin{cases} -1 \text{ if } c(y, x_j) - \phi_j < c(y, x_{j'}) - \phi_{j'} \quad \forall j' \neq j \\ 0 \text{ if } c(y, x_j) - \phi_j > \phi^c(y) \\ \text{not defined if } c(y, x_j) - \phi_j = \phi^c(y) \end{cases} \tag{4.1}$$

Using this equation we can show that

$$\frac{\delta C_{DP}}{\delta \phi_j} = a_j - \int_{V_{\phi_j}} \rho(y) dy \tag{4.2}$$

We will have a maximum when the derivative is equal to zero, assigning the same probability a_j to the Voronoi Power cell of $\nu(y)$. To find the exact value ϕ^* that maximizes the dual cost C_{DP} we can use a gradient ascend algorithm,

$$\phi_j^{k+1} = \phi_j^k + \Delta t \frac{\delta C_{DP}}{\delta \phi_j} \tag{4.3}$$

This approach can be refined and Newton methods for the calculation of Optimal the Voronoi Power Diagram have been proposed in the literature [32], [40].

4.2 Relation with Coverage Control

We start recalling the Coverage Control problem introduced in Chapter 2 and we will relate it to the Wasserstein distance. Our approach is inspired by the analysis of [28], where the relationship between Coverage Control and Optimal Transport was also discussed. In the Coverage problem we consider n sensors with positions $P = (p_1, \dots, p_n)$, $p_i \in \Omega$ and we maximize the sensing of a density $\rho(x)$, that represents the probability that some event occurs. Given the position of the sensors $P = (p_1, \dots, p_n)$ and a partition $\mathcal{W} = \{W_1, \dots, W_n\}$ of Ω we define the coverage cost as

$$\mathcal{H}(P, \mathcal{W}) = \sum_{i=1}^{n} \int_{W_i} ||x - p_i||^2 \rho(x) dx \tag{4.4}$$

It can be proved that the Voronoi Partition (1) gives the optimal partition for a set of points $P = (p_1, \dots, p_n)$.

$$\mathcal{H}(P) = \sum_{i=1}^{n} \int_{V_i} ||x - p_i||^2 \rho(x) dx$$

We note that the coverage cost (4.4) is very similar to the Wasserstein-2 distance (4.5) between two densities, both exhibit the quadratic term $||x - y||^2$

$$W_2^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} ||x - y||^2 d\pi(x, y) \tag{4.5}$$

Since $\mu = \sum_{i=1}^{n} a_i \delta_{x_i}(x)$ is discrete and $\nu(y) = \rho(y)dy$ is continuous we know from (6) that there will exist a map $T^* : \Omega \to \Omega$ such that $\pi^* = (T^*, id)_{\#\nu}$, and since T^* is single-valued every transport plan $\pi(x, y)$ will have the form

$$\pi(x, y) = \sum_{i=1}^{n} a_i \delta_{x_i}(x) \times \rho(y) \mathbb{1}_{W_i^*} dy$$

where $\mathbb{1}_{W_i^*}$ is the indicator function of the set $W_i^* = T^{*-1}(x_i)$ and $\mathcal{W}^* = (W_1^*, \dots, W_n^*)$ is a partition of $spt(\nu) = \Omega$. Additionally the mass is conserved, and we have $a_i = \nu(W_i^*) = \int_{W_i^*} \rho(y)dy$. Taking into account this considerations we can rewrite (4.5) as

$$W_2^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} ||x - y||^2 \sum_{i=1}^{n} a_i \delta_{x_i}(x) dx \times \rho(y) \mathbb{1}_{W_i^*} dy =$$

$$W_2^2(\mu, \nu) = \inf_{\substack{(W_1,\dots,W_n) \\ \nu(W_i) = a_i}} \sum_{i=1}^{n} \int_{W_i} ||x_i - y||^2 \rho(y) dy \tag{4.6}$$

In the coverage problem, if we consider P constant our objective is

$$\min_{\mathcal{W}} \mathcal{H}(P, \mathcal{W}) = \min_{(W_1,\dots,W_n)} \sum_{i=1}^{n} \int_{W_i} ||x - p_i||^2 \rho(x) dx$$

The objective function of coverage cost (4.4) and the Wasserstein distance (4.6) is the same, but we can see that when choosing the partition $\mathcal{W}$ in the Optimal Transport problem we need to impose that the mass is conserved in $\nu(W_i) = a_i$ while in the coverage problem we don't impose this constrain. Additionally the optimal partition for the Coverage problem is given by the Voronoi Partition $\mathcal{V} = (V_1, \dots, V_n)$, with this considerations we can establish that the coverage cost is a relaxation of the Optimal Transport cost

$$\min_{(a_1,\ldots,a_n)\in\mathbb{R}^n_{\geq 0}} \inf_{\substack{(W_1,\ldots,W_n)\\ \nu(W_i)=a_i}} W_2^2\left(\sum_{i=1}^n a_i\delta_{x_i}(x),\nu\right) = \min_{(W_1,\ldots,W_n)} \mathcal{H}(P,\mathcal{W})$$

And the minimum will be achieved when $a_i = \nu(V_i)$ for all i. If the weights $a_i \neq \nu(V_i)$, we have discussed at the beginning of the chapter that the optimal partition of the Optimal Transport problem will be given by a Voronoi Power cell. We note that Voronoi cells and Voronoi Power cells have a very similar structure. In the literature this considerations have been used to develop a generalization of Lloyd algorithm [10] that take into account the fixed mass constrains. This results were presented in the context of Quantization Theory, the study of discretization of continuous signals in space. The ideas of capacity constrains was also implemented in [44], in which capacity constrains were added to Coverage Control.

Remark 3. *It is important to note that while Lloyd's algorithm is a gradient flow of the Coverage cost $\mathcal{H}(P)$ it is not a gradient flow of Optimal Transport $W_2^2(\mu,\nu)$. In Lloyd's algorithm the set of admissible measures is always discrete, while in optimal transport we allow the measures to be continuous. To see this we will consider a discrete measure $\mu(y) = \sum_{i=1}^n a_i\delta_{x_i}(x)$ and a continuous measure $\nu(y) = \rho(y)dy$. Since ν is continuous there exists an optimal transport map $T_{\nu\to\mu} : Y \to X$ and the geodesic between ν and μ is given by*

$$\nu_t = (1-t)y + tT_{\nu\to\mu}(y) \quad y \in spt(\nu)$$

And the measures ν_t will be continuous for $t \in [0,1)$.

4.3 The continuous limit

Our next objective will be to study the continuous limit $n \to \infty$ of the Coverage Control algorithm, we will follow the developments of [28]. We will consider n agents with positions samples following $x_i \sim \mu$, with μ a continuous distribution. From

$$\mathcal{H}(x_1,\ldots,x_n) = W_2^2(\sum_{i=1}^n \nu(V_i)\delta_{x_i}(x),\nu)$$

We can consider the measure

$$\hat{\mu}_c(x) = \sum_{i=1}^n \nu(V_i)\delta_{x_i}(x)$$

and set $n \to \infty$. If $spt(\nu) \subset spt(\mu)$ applying the Glivenko Cantelli Theorem [4], the measure $\hat{\mu}_c(x) = \sum_{i=1}^n \nu(V_i)\delta_{x_i}(x)$ will converge weakly almost surely to the measure ν,

$$\lim_{n\to\infty} \sum_{i=1}^n \nu(V_i)\delta_{x_i}(x) = \nu$$

In this situation, considering $x_i \sim \mu$, the Coverage cost between the density μ and the density ν will be given by

$$\lim_{n\to\infty} \mathcal{H}(x_1,\ldots,x_n) = \lim_{n\to\infty} W_2^2\left(\sum_{i=1}^{n} \nu(V_i)\delta_{x_i}(x), \nu\right) = W_2^2(\nu,\nu) = 0$$

This result tells us that Coverage Control cannot be analyzed macroscopically, since every measure μ such that $spt(\nu) \subset spt(\mu)$ will give us a null coverage cost.

In [28] they proposed to consider a set of points $x_i \sim \mu$ and define the empirical probability measure

$$\hat{\mu}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}(x)$$

When we consider $n \to \infty$ we will have $\hat{\mu}_n \to \mu$ almost everywhere, from the Glivenko-Cantelli theorem [4]. Additionally,

$$\lim_{n\to\infty} W_2^2(\hat{\mu}_n,\nu) \le \lim_{n\to\infty} W_2^2(\hat{\mu}_n,\mu) + W_2^2(\mu,\nu) = W_2^2(\mu,\nu)$$

In next chapter we will use this model to formulate a Multi-Agent transport model.

Chapter 5

A Multi-agent Optimal Transport algorithm

In this section we will study the problem of transportation of large swarms of robots, different approaches have been proposed in the literature to treat this problem, some of them include Markov Transition matrices [5], [7], [16] and continuum models [29], [22]. We will focus on continuum models, where the swarm can be abstracted as a fluid and we study how the density of the agents evolve. The study of continuum models allows us to use tools like Lyapunov analysis and PDE analysis in order to prove convergence of the algorithms.
Recent work [30], [28] has studied the application of Optimal Transport to develop Multi-Agent transport algorithms. In this chapter we will focus on their approach and study the algorithms they proposed. Their approach to the Multi-Agent Transport problem consists in an iterative scheme that updates the position of the agents following a proximal point algorithm. It can be proven that this iterative scheme is equivalent to a gradient flow of the Wasserstein distance and the swarm will evolve following the Wasserstein geodesic. To implement the algorithm they rely on the Kantorovich potential, which can be estimated following a primal-dual algorithm.
We have joined the project and we have provided small modifications in the theory that improve the convergence of the algorithm. We have also observed that the algorithm that was presented can be used with non-convex domain and we have provided some simulations in non-convex domain. Finally we have started studying possible ways to implement collision avoidance in the movement of the swarm but our work is still in its early stages.

5.1 Problem definition

We start considering n identical agents with sensing and communicating capabilities located at positions $\boldsymbol{x} = (x_1, \ldots, x_n)$, which are sampled $x_i \sim \mu$ from a continuous probability distribution. We will consider the empirical distribution of the agents, given by

$$\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}(x)$$

When $n \to \infty$ the Glivenko-Cantelli theorem [4] guarantees that the discrete probability measure coverges $\hat{\mu} \to \mu$ almost everywhere. We will develop our algorithm with the discrete measure $\hat{\mu}$ and analyze it using the continuous model μ. This approximation will be sound when n is large. Our objective will be to efficiently transport μ into μ^*, a target probability measure, that represents the demand of agents in the space. We will assume μ^* continuous. We will minimize the transport cost between μ and μ^* according to a cost function $c : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ that satisfies the following assumption.

Assumption 1. *We assume that the cost function is a Riemannian metric and is conformal to the Euclidean distance with a strictly positive conformal factor $\xi(x) \in C^1(\Omega)$ and*

$$c(x, y) = \inf_{\substack{\gamma(t):[0,1]\to\Omega \\ \gamma(0)=x,\, \gamma(1)=y}} \int_0^1 \xi(\gamma(t))||\dot{\gamma}(t)||dt$$

We note that this cost is defined as a geodesic cost, and as such is well defined for path-connected non-convex domains. We would also like to point out that most of the results that we have studied in Chapter 3 still hold for a Riemannian metric, for a detailed treatment we refer the reader to [53].
Finally, it is important to note that there may be cases in which the cost that we are considering doesn't satisfy the Twist condition and we cannot prove the existence of an optimal transport map T^* that minimizes the Monge problem. We will provide an example of one of such cases in Annex A of how the Twist condition may be violated in a non-convex domain. To bypass this problems we will rely on rely on (6)

Proposition 16. *Let Ω be compact, $\mu, \nu \in \mathcal{P}(\Omega)$ and $c : \Omega \times \Omega \to \mathbb{R}$ continuous. Then the Kantorovich problem admits a solution. Moreover, if the measure μ is atom-less (i.e $\mu(\{x\}) = 0$, $\forall x \in X$), we have*

$$\min(KP) = \inf_{T_{\#\mu}=\nu} \int_X c(x, T(x))d\mu(x)$$

Despite not knowing if there exist an actual minimizer T^* of the Monge problem we can consider a transport map T that achieves a cost arbitrarily close to the infimum cost. We will call T_ϵ such that $\int_X c(x, T_\epsilon(x))d\mu(x) < \min(KP) + \epsilon$ an ϵ-sub-optimal transport map. In the rest of the chapter we will only talk about optimal transport maps T^*, making the abuse of notation $T^* = T_\epsilon$, for some $0 < \epsilon \ll 1$ when an optimal transport map doesn't exist.

5.2 An iterative transport method

In order to transport the source probability density μ into μ^* we will study the optimal transport geodesics, which will be parametric curves μ_t with $\mu_0 = \mu$ and $\mu_T = \mu^*$ that minimize the Wasserstein distance. We will discretize the Wasserstein distance geodesic into a sequence of points $\{\mu_k\}_{k=1,\dots,K}$ with $\mu_0 = \mu$ and $\mu_K = \mu^*$. From the triangular inequality in the Wasserstein space, we know that the intermediate points μ_k will belong to a geodesic if and only if

$$W_p(\mu, \mu^*) = \sum_{k=1}^{K} W_p(\mu_{k-1}, \mu_k)$$

From this we can generate the sequence $\{\mu_k\}_{k=1,\ldots,K}$ following

$$\begin{aligned} \mu_{k+1} \in \arg \min_{\rho \in \mathcal{P}(\Omega)} \quad & W_p(\mu_k, \rho) + W_p(\rho, \mu^*) \\ s.t \quad & W_p(\mu_k, \rho) \leq \epsilon \end{aligned} \tag{5.1}$$

We can evaluate this scheme in every point of the domain, giving us

$$\begin{aligned} x(k+1) \in \arg\min_{z \in \Omega} \quad & c(x(k), z) + c(z, T^*(x(k))) \\ s.t \quad & c(x(k), z) \leq \epsilon \end{aligned} \tag{5.2}$$

Likewise, the law of (5.2) with $x(0) \sim \mu$ evolves as (5.1). This equivalence allows us to update the positions in Ω and lift the update to the probability space. Now we introduce the process

$$\begin{aligned} x(k+1) \in \arg\min_{z \in \Omega} \quad & c(x(k), z) + \phi_{\mu_k \to \mu^*}(z) \\ s.t \quad & c(x(k), z) \leq \epsilon \end{aligned} \tag{5.3}$$

In the next proposition we will show that this new iterative scheme (5.3), which formulates the problem in the Kantorovich formulation, is equivalent to the iterative scheme (5.2), that formulates the problem in the Monge formulation. This result will be fundamental, allowing us to develop an Optimal Transport algorithm without having to calculate the transport map $T^*(x)$.

Proposition 17. *The process (5.2) and (5.3) are equivalent. The equivalence is in the sense that the set of minimizers are equal.*

Proof. We start with the c-conjugate property of the Kantorovich potential (7),

$$\phi_{\mu_k \to \mu^*}(x) = \inf_z c(x, z) - \phi^c_{\mu_k \to \mu^*}(z)$$

From (1) we know that $\phi^c_{\mu_k \to \mu^*}(z) = -\phi_{\mu_k \to \mu^*}(z)$ and,

$$\phi(x) = \inf_z c(x, z) - \phi_{\mu_k \to \mu^*}(z)$$

Then,

$$\phi_{\mu_k \to \mu^*}(x) \leq c(x, z) - \phi_{\mu_k \to \mu^*}(z) \tag{5.4}$$

Taking $z = T_k(x)$ and integrating we have

$$\int_\Omega (\phi_{\mu_k \to \mu^*}(x) - \phi_{\mu_k \to \mu^*}(T_k(x))) d\mu_k(x) = \int_\Omega \phi_{\mu_k \to \mu^*}(x) d\mu_k(x) - \int_\Omega \phi_{\mu_k \to \mu^*}(y) d\mu^*(y) \leq \int_\Omega c(x, T_k(x)) d\mu_k(x)$$

Where the RHS of the inequality corresponds to the Monge Optimal Transport cost and the LHS of the inequality corresponds to the Kantorovich Transport cost. From (6) we know

that the cost of both formulations is equal.
This gives us the equality

$$\phi_{\mu_k \to \mu^*}(x) - \phi_{\mu_k \to \mu^*}(T_k(x)) = c(x, T_k(x))$$

We now observe that $c(x, T_k(x) = c(x, z) + c(z, T_k(x))$ if and only if z belongs to the geodesic between x and $T_k(x)$. Taking z in the geodesic we can write

$$\phi_{\mu_k \to \mu^*}(x) + \phi_{\mu_k \to \mu^*}(z) - \phi_{\mu_k \to \mu^*}(z) - \phi_{\mu_k \to \mu^*}(T_k(x)) = c(x, z) + c(z, T_k(x))$$

Rearranging terms,

$$[\phi_{\mu_k \to \mu^*}(x) - \phi_{\mu_k \to \mu^*}(z) - c(x, z)] + [\phi_{\mu_k \to \mu^*}(z) - \phi_{\mu_k \to \mu^*}(T_k(x)) - c(z, T_k(x))] = 0$$

We know from (5.4) that each of the terms is non-negative, so they must be 0. This gives us $\phi_{\mu_k \to \mu^*}(x) = c(x, z) + \phi_{\mu_k \to \mu^*}(z)$. Then if $c(x, z) \leq \epsilon$ z is a minimizer of (5.3). Since z can only be choosen as a point that belongs to the geodesic it is also a minimizer of (5.2), proving the equivalence of the minimizers of (5.3) and (5.2). □

5.2.1 Estimating the Kantorovich Potential

In order to be able to implement the update scheme (5.3) we need to know the Kantorovich potential. In this section we will construct a primal-dual algorithm that will allow us to estimate the Kantorovich potential $\phi_{\mu_k \to \mu^*}$.
We start with the Kantorovich dual formulation of Optimal Transport, which we recall from (3.10),

$$\begin{aligned} C_{DP} = \sup_{\phi} \quad & \int_\Omega \phi(x) d\mu(x) - \int_\Omega \phi(y) d\mu^*(y) \\ \text{s.t.} \quad & |\phi(x) - \phi(y)| \leq c(x, y) \end{aligned} \tag{5.5}$$

In the multi-agent transport problem we only have n agents with positions $\boldsymbol{x} = (x_1, \ldots, x_n)$ that make up the support of $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}(x)$, then

$$\int_\Omega \phi(x) \frac{1}{n} \sum_{i=1}^n \delta_{x_i}(x) dx = \sum_{i=1}^n \frac{1}{n} \phi^i$$

The Kantorovich potential can be estimated as $\Phi : \mathbb{N} \times \Omega \to \mathbb{R}$ generated by the values $(\phi_1(k), \ldots, \phi_n(k))$ defined at time $t = k$ on the positions of the agents, $\Phi(k, x_i(k)) = \phi^i$. To extend this definition to the rest of the points in Ω we can use a piece-wise constant interpolation generated by the Voronoi Partition (1) of the domain Ω with respect to the cost, $V = (V_1, \ldots, V_n)$. Then $\Phi(k, x) = \phi^i \quad \forall x \in V_i$.
The Voronoi partition allows us to define a graph $G = (\{x_i\}_{i=1,\ldots,n}, E)$ where we say we have an edge $(i, j) \in E$ if $V_i \cap V_j \neq \emptyset$. To obtain a decentralized and efficient algorithm we will only impose the constrain $|\phi(x_i) - \phi(x_j)| \leq c(x_i, x_j)$ if $(i, j) \in E$.

Taking into account these considerations we can rewrite the problem as

$$\begin{aligned} C_{DP} = \max_{(\phi_1,\ldots,\phi_n)} \quad & \sum_{i=1}^{n} \phi^i \frac{1}{n} - \phi^i \mu^*(V_i) \\ \text{s.t.} \quad & |\phi^i - \phi^j| \leq c(x_i, x_j) \quad \forall (i,j) \in E \end{aligned} \tag{5.6}$$

Where $\mu^*(V_i) = \int_{V_i} d\mu^*(x)$. This is a linear programming problem and it can be solved using a primal-dual method. The first step is to define the Lagrangian of (5.6),

$$L(\phi, \lambda) = \sum_{i=1}^{N} \kappa \phi^i \left(\frac{1}{N} - \mu^*(V_i) \right) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j \in N_i} \lambda_{ij} \left(|\phi^i - \phi^j|^2 - c_{ij}^2 \right) \tag{5.7}$$

Where κ is a constant that we will denote as gain and $\lambda \geq 0$ is a Lagrange Multiplier. From the Lagrangian we can write a primal ascend, dual descend algorithm using the dynamics

$$\begin{aligned} \phi^i(l+1) &= \phi^i(l) + \frac{\delta L}{\delta \phi^i}(\phi, \lambda) \\ \lambda_{ij}(l+1) &= \max \left\{ 0, \lambda_{ij}(l) - \frac{\delta L}{\delta \lambda_{ij}}(\phi, \lambda) \right\} \end{aligned}$$

Which results in

$$\begin{aligned} \phi^i(l+1) &= \phi^i(l) + \tau \left(\kappa \left(\frac{1}{N} - \mu^*(V_i) \right) - \sum_{j \in N_i} \lambda_{ij}(l)(\phi^i(l) - \phi^j(l)) \right) \\ \lambda_{ij}(l+1) &= \max \left\{ 0, \lambda_{ij}(l) + \frac{\tau}{2} \left(|\phi^i - \phi^j|^2 - c_{ij}^2 \right) \right\} \quad \text{with } j \in N_i \end{aligned} \tag{5.8}$$

The algorithm is decentralized because it only needs nearest neighbours information to perform the updates. In fact we can write λ as a weighted Laplacian of the Voronoi graph $G = (\{x_i\}, E)$, $(x_i, x_j) \in E$ if $\delta V_i \cap \delta V_j \neq \emptyset$, with λ_{ij} the weight of the edges.
The model that we have described is summarized in Algorithm 5.2.1. In the algorithm the primal-dual method is applied during n steps before every update without waiting for convergence.

5.3 Analysis of PDE model

In this section we will analyze the behaviour of Algorithm 5.2.1 for continuous time and $n \to \infty$. This will allow us to use partial differential equations to study the behaviour of the system.
We will start introducing some technical results that we will later use to prove the convergence. Then we will write an alternative Continuous formulation of the Kantorovich Dual Formulation and we will use this formulation to prove the convergence of the primal-dual algorithm (5.8) in the continuum.
We will provide a continuous formulation of the transport update scheme (5.3). We will use

Algorithm 1 Multi-agent (on-the-fly) optimal transport

Input: Target measure μ^*, transport cost $c(x,y)$, Bound on step size ϵ, time step τ.
For each agent i at time instant k of transport:

1: Obtain positions $x_j(k)$ of neighbours within a communication radius r large enough to cover all Voronoi neighbours.
2: Compute Voronoi cell V_i, the mass of the Voronoi cell $\mu^*(V_i)$ and the Voronoi neighbour N_i
3: Initialize $\phi^i \leftarrow \Phi(k-1, x_i(k))$, $\lambda_{ij} \leftarrow \lambda_{ij}(k-1)$.
4: Implement n iterations of the primal-dual algorithm (5.8) synchronously in communication with $j \in N_i$ to obtain $\phi^i(k)$, $\lambda_{ij}(k)$
5: Communicate with neighbors $j \in N_i$ to obtain $\phi^j(k)$ and construct a local estimate of $\Phi(k,x)$ with multivariate interpolation.
6: Implement transport step (5.3) with the local estimate $\Phi(k,x)$, which approximates $\phi_{\mu_k \to \mu^*}$

this formulation to prove exponential convergence of the algorithm in the continuum. We will also discuss the convergence of the on-the-fly algorithm, where we implement the estimation of the Kantorovich potential and the transport at the same time and consider a constant Lagrange Multiplier $\lambda = \lambda(x) > 0$.

5.3.1 Technical helpers

We start introducing some results that will help us prove the convergence of the algorithm.

Lemma 2 (Rellich-Kondrachov Compactness Theorem)**.** *Let $\Omega \subset \mathbb{R}^N$ be a open, bounded and such that $\delta\Omega$ is C^1. Suppose $1 \leq p < n$ then the Sovolev space $W^{1,p}(\Omega)$ is compactly embedded in $L^q(\Omega)$ for each $1 \leq q < \frac{pn}{n-p}$. In particular, we have $W^{1,p}(\Omega)$ is compactly contained in $L^p(\Omega)$.*

Lemma 3 (LaSalle Invariance Principle)**.** *Let $\{\mathcal{P}(t) | t \in \mathbb{R}_{\geq 0}\}$ be a continuous semigroup of operators on a Banach space U (closed subset of a Banach space with norm $||\cdot||$), and for any $u \in U$, define the positive orbit starting for u at $t=0$ as $\Gamma_+(u) = \{\mathcal{P}(t)u | t \in \mathbb{R}_{\geq 0}\} \subseteq U$. Let $V : U \to \mathbb{R}$ be a continuous Lyapunov functional on $\mathcal{G} \subset U$ for $\mathcal{P}$ (such that $\dot{V}(u) = \frac{d}{dt}V(\mathcal{P}(t)u) \leq 0$ in $\mathcal{G}$). Define $E = \{u \in \bar{\mathcal{G}} | \dot{V}(u) = 0\}$ and let $\hat{E}$ be the largest invariant subset of E. If for $u_0 \in \mathcal{G}$, the orbit $\Gamma_+(u_0)$ is pre-compact (lies in a compact subset of U), then $\lim_{t\to+\infty} d_U(\mathcal{P}(t)u_0, \hat{E}) = 0$, where $d_U(y, \hat{E}) = \inf_{x \in \hat{E}} ||y - x||_U$, where d_U is the distance in U.*

Lemma 4 (Divergence Theorem)**.** *For a smooth vector field $\boldsymbol{F}$ over a bounded open set $\Omega \subset \mathbb{R}^N$ with boundary $\delta\Omega$, the volume integral of the divergence $\nabla \cdot \boldsymbol{F}$ of $\boldsymbol{F}$ over Ω is equal to the surface integral of $\boldsymbol{F}$ over $\delta\Omega$:*

$$\int_\Omega (\nabla \cdot \boldsymbol{F}) d\mu = \int_{\delta\Omega} \boldsymbol{F} \cdot \boldsymbol{n} dS$$

where $\boldsymbol{n}$ is the outward normal to the boundary and dS the measure on the boundary. For a scalar field ψ and a vector field $\boldsymbol{F}$ defined over $\Omega \subset \mathbb{R}^n$:

$$\int_\Omega (\boldsymbol{F} \cdot \nabla\psi) d\mu = \int_{\delta\Omega} \psi(\boldsymbol{F} \cdot \boldsymbol{n}) dS - \int_\Omega \psi(\nabla \cdot \boldsymbol{F}) d\mu$$

5.3.2 A continuous formulation of the Kantorovich problem

Here we recall the theory that we have studied in Chapter 3 for costs that are a metric to provide an alternative formulation to (5.5). We recall from Chapter 3 the following proposition (11). We provide the proof again and note that we only require the domain Ω to be path-connected, the domain may be non-convex.

Proposition 18. *The constrain $|\phi(x) - \phi(y)| \leq c(x, y)$ is equivalent to $||\nabla\phi(x)|| \leq \xi(x)$*

Proof. Let $x \in \mathring{\Omega}$ and $y \in B(x, \epsilon) \subset \mathring{\Omega}$. Using the Mean Value Theorem, there exist $z_1 = m_1 x + (1 - m_1)y$, $z_2 = m_2 x + (1 - m_2)y$ with $m_1, m_2 \in [0, 1]$ such that

$$\frac{|\nabla\phi(z_1) \cdot (x - y)|}{||x - y||} = \frac{|\phi(x) - \phi(y)|}{||x - y||} \leq \frac{c(x, y)}{||x - y||} \leq \frac{\xi(z_2)||x - y||}{||x - y||}$$

If we take $y = x + tv$ and take $t \to 0$ then $z_1, z_2 \to x$ and we will obtain

$$\frac{\nabla\phi(x) \cdot v}{||v||} \leq \xi(x)$$

For every $v \in \mathcal{T}_x(\Omega)$, the tangent space at point $x \in \mathring{\Omega}$. This gives us the inequality $||\nabla\phi|| \leq \xi(x)$.
To prove the converse we assume $\nabla\phi \leq \xi(x)$ and we consider the geodesic γ of the cost c that connects $\gamma(0) = x$ and $\gamma(1) = y$. Giving us

$$|\phi(x) - \phi(y)| = \int_0^1 \nabla\phi \cdot \dot{\gamma}(t) dt \leq \int_0^1 ||\nabla\phi|| \, ||\dot{\gamma}(t)|| dt \leq \int_0^1 \xi(\gamma(t))||\dot{\gamma}(t)|| dt = c(x, y)$$

□

We also note that since the cost is C^1 from (10) we know that $\nabla\phi = \nabla_x c(x_0, y_0)$ $\forall (x_0, y_0) \in spt(\pi^*)$ and we can deduce that the Kantorovich potential is C^1 almost everywhere. Additionally, if Ω is path-connected the equality of the gradients forces the uniqueness up to additive constants of the Kantorovich potential, as we have discussed in (14). With this considerations we can rewrite the Kantorovich dual formulation as

$$\begin{aligned} C_{DP} = \sup_\phi \quad & \int_\Omega \phi(x) d\mu(x) - \int_\Omega \phi(y) d\mu^*(y) \\ \text{s.t.} \quad & |\nabla\phi(x)| \leq \xi(x) \end{aligned} \tag{5.9}$$

5.3.3 Convergence of the Primal-Dual Flow

We will see how we can use (5.9) to prove the convergence of the primal-dual scheme (5.8) in continuous time and space (the limit $n \to \infty$). The proof that we present is extracted from [30] but we have made minor modifications.

We start by defining the Lagrangian of (5.9), which corresponds to the discrete time and discrete space Lagrangian (5.7).

$$L(\phi, \lambda) = \int_\Omega \kappa\phi(\rho - \rho^*) - \frac{1}{2}\int_\Omega \lambda(|\nabla\phi|^2 - |\xi|^2) \tag{5.10}$$

Where we integrate with respect to the Lebesgue measure and κ is a constant. From the Lagrangian we can obtain the optimality conditions.

Lemma 5. *The necessary and sufficient conditions for a feasible solution $\bar{\phi}$ of (5.9) to be optimal are:*

$$\begin{aligned} -\nabla\cdot(\bar{\lambda}\nabla\bar{\phi}) = \kappa(\rho-\rho^*) \quad &(\textit{in } \Omega)\\ \bar{\lambda}\nabla\bar{\phi}\cdot\boldsymbol{n} = 0 \quad &(\textit{on } \delta\Omega)\\ \bar{\lambda}\geq 0, \quad |\nabla\bar{\phi}|\leq\xi \quad &(\textit{Feasibility})\\ \bar{\lambda}(|\nabla\bar{\phi}|-\xi) = 0\, a.e. \quad &(\textit{Complementary slackness}) \end{aligned} \tag{5.11}$$

Proof. We consider the Lagrangian (5.10),

$$L(\phi, \lambda) = \int_\Omega \kappa\phi(\rho - \rho^*) - \frac{1}{2}\int_\Omega \lambda(|\nabla\phi|^2 - |\xi|^2)$$

We will now consider the first variation of the functional with respect to the variation $\delta\phi$, giving us

$$\left\langle \frac{\delta L}{\delta\phi}, \delta\phi \right\rangle = \int_\Omega \kappa(\rho-\rho^*)\delta\phi - \int_\Omega \lambda\nabla\phi\cdot\nabla\delta\phi$$

$$= \int_\Omega \kappa(\rho-\rho^*)\delta\phi + \int_\Omega \nabla\cdot(\lambda\nabla\phi)\delta\phi - \int_{\delta\Omega}\lambda\nabla\phi\cdot\boldsymbol{n}\delta\phi$$

where the last equality comes from the divergence theorem (4). If a solution $(\bar{\phi}, \bar{\lambda})$ is a stationary point of the $L(\phi,\lambda)$ then $\left\langle \frac{\delta L}{\delta\phi}, \delta\phi \right\rangle = 0$ for any variation $\delta\phi$ around the stationary point $(\bar{\phi}, \bar{\lambda})$. This gives us $-\nabla\cdot(\bar{\lambda}\nabla\bar{\phi}) = \kappa(\rho-\rho^*)$ in Ω and $\bar{\lambda}\nabla\bar{\phi}\cdot\boldsymbol{n} = 0$ on $\delta\Omega$.
We remember that the objective function $\int_\Omega \phi(x)d\mu(x) - \int_\Omega \phi(y)d\mu^*(y)$ is linear and the constrains $|\nabla\phi(x)| \leq \xi(x)$ are convex. We can the apply the Karush Kuhn Tucker condition to obtain the necessary and suficient conditions for optimality. The feasibility condition for Lagrange multiplier is $\bar{\lambda} \geq 0$ and $|\nabla\bar{\phi}| \leq \xi$ is the feasibility condition of ϕ, the primal variable. Finally $\bar{\lambda}(|\nabla\bar{\phi}|-\xi) = 0\, a.e.$ is the complementary slackness condition. □

From the optimality conditions we can define a primal-dual flow that converges to the saddle-point $(\bar{\phi}, \bar{\lambda})$ of the Lagrangian (5.10).

$$\begin{aligned} \delta_t \phi &= \nabla \cdot (\lambda \nabla \phi) + \kappa(\rho - \rho^*) \\ \nabla \phi \cdot \boldsymbol{n} &= 0 \quad (\text{on } \delta\Omega) \\ \delta_t \lambda &= \frac{1}{2}\left[|\nabla \bar{\phi}|^2 - \xi^2)\right]^+_\lambda \\ \phi(0,x) &= \phi_0(x) \quad \lambda(0,x) = \lambda_0(x) \end{aligned} \tag{5.12}$$

where $[f]^+_\lambda = \begin{cases} f & \text{if } \lambda > 0 \\ \max\{0, f\} & \text{if } \lambda = 0 \end{cases}$ is a projection operator. The PDE model corresponds to the dynamics $\delta_t \phi = \frac{\delta L}{\delta \phi}$ and $\delta_t \lambda = -\frac{\delta L}{\delta \lambda}$. In [30] the existence and uniqueness of solutions of this PDE is discussed. We now present an assumption on the well-posedness of the primal-dual flow.

Assumption 2. *We assume that (5.12) is well posed, with solution (ϕ, λ) such that $\phi \in L^\infty(0, \infty; H^1(\Omega))$, the Lagrange multiplier function $\lambda \in L^\infty(0, \infty; L^\infty(\Omega))$ and is precompact in $L^2(\Omega)$.*

This leads to the following lemma on the convergence of the primal-dual flow.

Lemma 6 (Convergence of Primal-Dual flow [30])**.** *The solutions (ϕ_t, λ_t) to the primal-dual flow (5.12), under Assumption 2 on the well-posedness of the primal-dual flow, converge to an optimizer $(\bar{\phi}, \bar{\lambda})$ given in (5.11) in the L^2 norm as $t \to \infty$, for any fixed $\rho, \rho^* \in L^2(\Omega)$.*

Proof. We start considering $(\bar{\phi}, \bar{\lambda})$ a solution of the opitimality conditions (5.11), we will consider a Lyapunov function $V(\phi, \lambda) = \frac{1}{2}\int_\Omega |\phi - \bar{\phi}|^2 dx + \frac{1}{2}\int_\Omega |\lambda - \bar{\lambda}|^2 dx$. We have $V(\phi, \lambda) > 0$ for all $\phi, \lambda \in L^2(\Omega)$. We will see that the time derivative of the Lyapunov function is non-increasing, $\dot{V} \leq 0$. Using that $\delta_t \phi = \frac{\delta L}{\delta \phi}$ and $\delta_t \lambda = \left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda$ we can write the time derivative of the Lyapunov function as

$$\dot{V} = \langle \frac{\delta L}{\delta \phi}, \phi - \bar{\phi} \rangle + \langle \left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda, \lambda - \bar{\lambda} \rangle =$$

$$\dot{V} = \langle \frac{\delta L}{\delta \phi}, \phi - \bar{\phi} \rangle - \langle \frac{\delta L}{\delta \lambda}, \lambda - \bar{\lambda} \rangle + \langle \frac{\delta L}{\delta \lambda} + \left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda, \lambda - \bar{\lambda} \rangle =$$

Since $L(\phi, \lambda)$ is concave in ϕ and convex in λ we get:

$$\dot{V} \leq L(\phi, \lambda) - L(\bar{\phi}, \lambda) + L(\phi, \bar{\lambda}) - L(\phi, \lambda) + \langle \frac{\delta L}{\delta \lambda} + \left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda, \lambda - \bar{\lambda} \rangle$$

$$\dot{V} \leq L(\bar{\phi}, \bar{\lambda}) - L(\bar{\phi}, \lambda) + L(\phi, \bar{\lambda}) - L(\bar{\phi}, \bar{\lambda}) + \langle \frac{\delta L}{\delta \lambda} + \left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda, \lambda - \bar{\lambda} \rangle$$

And since $(\bar{\phi}, \bar{\lambda})$ is a saddle point of $L(\phi, \lambda)$ we have that $L(\bar{\phi}, \bar{\lambda}) - L(\bar{\phi}, \lambda) \leq 0$ and $L(\phi, \bar{\lambda}) - L(\bar{\phi}, \bar{\lambda}) \leq 0$. Moreover, by definition, when $\lambda(t,x) > 0$ we have $\left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda = -\frac{\delta L}{\delta \lambda}$ and when $\lambda(t,x) = 0$ $\lambda - \bar{\lambda} \leq 0$ and $\left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda \geq -\frac{\delta L}{\delta \lambda}$ at (t,x). This leads to $\langle \frac{\delta L}{\delta \lambda} + \left[-\frac{\delta L}{\delta \lambda}\right]^+_\lambda, \lambda - \bar{\lambda} \rangle \leq 0$

at any (t,x), therefore $\dot{V} \leq 0$. Moreover, by Assumption (2) the problem is well posed and the orbit ϕ is bounded in $H^1(\Omega)$ which, by Lemma 2 it is compactly embedded in $L^2(\Omega)$. Moreover, by Assumption 2 we have that λ is pre-compact in $L^2(\Omega)$. Additionally, $\dot{V} = 0$ only at an optimizer $(\bar{\phi}, \bar{\lambda})$, which implies that the flow converges asymptotically to solution $(\bar{\phi}, \bar{\lambda})$ of the optimality conditions (5.11). □

5.3.4 The transport equation

After studying the estimation of the Kantorovich potential in the continuous we will now study the transport in continuous time. To do it we will consider the iteration scheme (5.3) and study the limit $\epsilon \to 0^+$, leading to a continuous dynamics, which will correspond to a curve in the Wasserstein space. From the theory presented in Chapter 3 the absolutely continuous curves in the Wasserstein space correspond to solutions of the Continuity equation (3.5).

$$\delta_t \mu_t + \nabla \cdot (\mu_t \boldsymbol{v_t}) = 0$$

We also proved that a gradient flow on the Wasserstein distance is generated by the vector field $\boldsymbol{v_t} = -\nabla\phi$. In this section we will see that the transport generated by the update scheme (5.3) is given by a gradient flow of the Wasserstein distance, where $\boldsymbol{v_t} = -\alpha\nabla\phi$, with $\alpha > 0$ a positive function on Ω. We start with the iterative scheme (5.3),

$$\begin{aligned} x(k+1) \in \arg\min_{z\in\Omega} \quad & c(x(k), z) + \phi_{\mu_k \to \mu^*}(z) \\ s.t \quad & c(x(k), z) \leq \epsilon \end{aligned}$$

If x^+ is a minimizer, we have

$$c(x, x^+) + \phi(x^+) \leq \phi(x)$$

We will now consider a discrete time step for this evolution $\Delta t = g(\epsilon) \to 0$, with $g : \mathbb{R} \to \mathbb{R}$ monotonically increasing. Since $\phi_{\mu\to\mu^*}$ is bounded and continuously differentiable we know that $\lim_{\epsilon\to 0} x^+ = x$ exists. We define $\boldsymbol{v}(x) = \lim_{\Delta T\to 0} \frac{x^+ - x}{\Delta t}$, and we have

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t} c(x, x^+) \leq \lim_{\Delta t\to 0} \frac{1}{\Delta t}(\phi_{\mu\to\mu^*}(x) - \phi_{\mu\to\mu^*}(x^+))$$

It follows that

$$\xi(x)|\boldsymbol{v}(x)| \leq -\nabla\phi_{\mu\to\mu^*}(x) \cdot \boldsymbol{v}(x)$$

And since we have $|\nabla\phi_{\mu\to\mu^*}| \leq \xi$ the inequality will only hold if

$$\boldsymbol{v} = -\alpha\nabla\phi_{\mu\to\mu^*} \tag{5.13}$$

With $\alpha = \alpha(x) > 0$ a non-negative function in Ω. We will see that the function α is related to the choice of g such that $\Delta t = g(\epsilon)$. Additionally we will show that we can choose a non-increasing function $g : \mathbb{R} \to \mathbb{R}$ such that $\alpha = \bar{\alpha}$, a fixed value.

Proposition 19. *The function $\alpha > 0$ is uniquely defined for every choice of $g : \mathbb{R} \to \mathbb{R}$ monotonically increasing function.*
Conversely let $\bar{\alpha} > 0$, then there exists a monotonically increasing function $\bar{g} : \mathbb{R} \to \mathbb{R}$ such

that $\bar{\boldsymbol{v}} = \lim_{\epsilon \to 0} \frac{x^+ - x}{\bar{g}(\epsilon)}$, *with* x^+ *the update given by the iterative scheme (5.3), exists and* $\bar{\boldsymbol{v}} = -\bar{\alpha}\nabla\phi$.

Proof. For the first implication we note that $\boldsymbol{v} = \lim_{\epsilon \to 0} \frac{x^+ - x}{g(\epsilon)}$ is uniquely defined and from (14) the Kantorovich potential is unique up to additive constants, giving us a unique α.
For the second implication we can use the first implication to get

$$\boldsymbol{v} = \lim_{\epsilon \to 0} \frac{x^+ - x}{g(\epsilon)} = -\alpha\nabla\phi$$

For some $g : \mathbb{R} \to \mathbb{R}$ monotonically increasing and $\alpha > 0$. If we consider $\bar{g}(\epsilon) = g(\epsilon)\frac{\alpha(x)}{\bar{\alpha}(x)}$, the function $\bar{g}(\epsilon)$ is monotonically increasing since $\alpha, \bar{\alpha} > 0$ and we obtain

$$\bar{\boldsymbol{v}} = \lim_{\epsilon \to 0} \frac{x^+ - x}{\bar{g}(\epsilon)} = \lim_{\epsilon \to 0} \frac{x^+ - x}{g(\epsilon)} \frac{\bar{\alpha}}{\alpha} = \bar{\alpha}\nabla\phi$$

□

We also note that $\boldsymbol{v}(t, x) = \lim_{\epsilon \to 0} \frac{x^+(t) - x(t)}{g(\epsilon)}$ defines a parametrization of the curve $(\rho_t)_t$ defined through $\delta_t\rho_t + \nabla \cdot (\rho_t v_t) = 0$. When we change the function $g : \mathbb{R} \to \mathbb{R}$ we are also changing the parametrization of the curve $(\rho_t)_t$ defined through $\boldsymbol{v}(t, x)$. The process of finding a $\bar{g} : \mathbb{R} \to \mathbb{R}$ such that $\bar{\boldsymbol{v}} = \lim_{\epsilon \to 0} \frac{x^+ - x}{\bar{g}(\epsilon)} = -\bar{\alpha}\nabla\phi$ is a reparametrization of the curve.

5.3.5 Convergence of transport

We will now study the convergence of the transport

$$\begin{aligned} \delta_t\rho &= -\nabla \cdot (\rho\boldsymbol{v}) \\ \boldsymbol{v} &= -\alpha\nabla\phi_{\mu \to \mu^*} \end{aligned} \tag{5.14}$$

In (17) we have discussed that the set of minimizers of (5.3) belong to the geodesic connecting μ and μ^*. Since (5.14) is the continuous limit of (5.3) the solution of the PDE will be the geodesic μ_t. Now we focus on the well-posedness of the problem.

Remark 4 (Existence and Uniqueness of solutions to the transport PDE)**.** *We refer the reader to [2] for a detailed treatment of existence and uniqueness results for the transport equation with transport vector fields that are well-posed.*

Assumption 3 (Well-posedness of gradient flow on optimal transport)**.** *We assume that* μ^* *is absolutely continuous with a density* ρ^* *in* $H^1(\Omega)$ *and spt*$(\mu_0) = \Omega$. *Further we assume that (3.5) is well-posed for the gradient flow on the optimal transport cost, with solution* $\rho \in L^\infty(0, \infty, H^1(\Omega))$.

With this assumption we can state the following convergence theorem, that we have adapted from [30], adding an exponential convergence rate κ.

Theorem 7. *Under Assumption 3 on the well-posedness of the gradient flow on the optimal transport cost and for absolutely continuous initial distributions* μ_0 *with spt*$(\mu_0) = \Omega$, *the* *solution to the transport 3.5 with transport vector $\boldsymbol{v} = -\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$, with $\lambda_{\mu\to\mu^*}$ and $\phi_{\mu\to\mu^*}$ the Kantorovich potential and the Lagrange multiplier for the transport $\mu \to \mu^*$, converges exponentially $||\rho - \rho^*||_{L^2(\Omega)} \propto e^{-\kappa t}$.*

Proof. From the optimality conditions of the primal-dual flow (5.11) we have that $\nabla \cdot (\lambda_{\mu\to\mu^*}\nabla\phi_{\mu\to\mu^*}) = \kappa(\rho^* - \rho)$, which implies $\delta_t\rho = -\nabla\cdot(\rho\boldsymbol{v}) = \nabla\cdot(\lambda_{\mu\to\mu^*}\nabla\phi_{\mu\to\mu^*}) = \kappa(\rho^* - \rho)$ when $\rho > 0$. Moreover we have that ρ_0 and ρ^* are strictly positive in Ω. Therefore, for any $t \in [0,\infty]$ and $x \in \mathring{\Omega}$, we have $\rho(t,x) > 0$. Consequently, since $\rho(t,x) > 0$, the transport vector field $\boldsymbol{v} = -\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$ is well-defined in Ω.
Let $V : L^2(\Omega) \to \mathbb{R}_{\geq 0}$ be defined by $V(\rho) = \frac{1}{2}\int_\Omega |\rho - \rho^*|^2 dx$, where ρ is the density function of the continuous probability measure μ. The time derivative $\dot{V}$, under the transport (3.5) by $\boldsymbol{v} = -\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$ is given by:

$$\dot{V} = \int_\Omega (\rho - \rho^*)\delta_t\rho = -\int_\Omega (\rho - \rho^*)\nabla\cdot(\rho\boldsymbol{v}) = \int_\Omega (\rho - \rho^*)\nabla\cdot(\lambda_{\mu\to\mu^*}\nabla\phi_{\mu\to\mu^*})$$

Then,

$$\dot{V} = -\int_\Omega \kappa|\rho - \rho^*|^2 = -2\kappa V$$

And V is a Lyapunov functional for the transport by the vector field $\boldsymbol{v} = -\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$. Moreover, by Assumption 3 the solution ρ is bounded in $H^1(\Omega)$, which by the Rellich-Kondrachov theorem 2 is embedded in $L^2(\Omega)$. We then infer that the solution ρ to the transport (3.5) by the vector field $\boldsymbol{v} = -\frac{\lambda_{\mu\to\mu^*}}{\rho}\nabla\phi_{\mu\to\mu^*}$ is pre-compact and using LaSalle Invariance Principle in Lemma 3 we know $\lim_{t\to\infty} ||\rho - \rho^*||_{L^2} = 0$. Moreover, from $\dot{V} = -2\kappa V$ we have $||\rho - \rho^*||^2_{L^2(\Omega)} \propto e^{-2\kappa t}$ and $||\rho - \rho^*||_{L^2(\Omega)} \propto e^{-\kappa t}$. □

Remark 5. *The exponential convergence of the algorithm permits the adaptation of the multi-agent transport algorithm to tracking scenarios where the target distributions evolve on a slower timescale.*

Finally we consider the case where we don't wait for the primal-dual flow to converge before transporting the density. In this case the transport and the estimation of the Kantorovich potential are performed jointly. We will show that we can consider a constant Lagrange multiplier $\lambda(x) > 0$ and prove convergence. This leads to the following dynamics

$$\begin{aligned} \delta_t\phi &= \nabla\cdot(\lambda\nabla\phi) + \kappa(\rho - \rho^*) \\ \nabla\phi\cdot\boldsymbol{n} &= 0 \quad \text{on } \delta\Omega \\ \lambda &= \lambda(x) > 0 \end{aligned} \tag{5.15}$$

In [30] the existence and uniqueness of solutions of (5.15) was discussed, we will assume existence and uniqueness of solutions.

Assumption 4. *We assume that μ^* is absolutely continuous with a density ρ^* in $H^1(\Omega)$ and $spt(\mu_0) = \Omega$. Further we assume that the primal flow (5.15) is well-posed and the transport (3.5) are well-posed, with solution ϕ and ρ such that $\phi \in L^\infty(0,\infty;H^1(\Omega))$, and strictly positive $\rho \in L^\infty(0,\infty;H^1(\Omega))$.*

Finally we present a convergence theorem on the convergence of (5.15) and prove it following [30].

Theorem 8 (Convergence of on-the-fly transport)**.** *Under Assumption 4, the solutions ρ to (3.5) with $\boldsymbol{v} = -\kappa\frac{\lambda}{\rho}\nabla\phi$, with ϕ from (5.15) converge in the L^2-norm to ρ^* as $t \to \infty$, while the solutions to the primal flow (5.15) converge to the optimality condition (5.11) corresponding to $\rho = \rho^*$.*

Proof. We note that from Assumption 4 $\rho > 0$ and the transport vector field $\boldsymbol{v} = -\kappa\frac{\lambda}{\rho}\nabla\phi$ is well-defined on Ω. We now consider the following Lyapunov functional:

$$E = \frac{1}{2}\int_\Omega \lambda|\nabla\phi|^2 + \frac{1}{2}\int_\Omega |\rho - \rho^*|^2$$

The time derivative of E under the flow (5.15) and $\boldsymbol{v} = -\kappa\frac{\lambda}{\rho}\nabla\phi$ is given by

$$\dot{E} = \int_\Omega \lambda\nabla\phi\cdot\nabla\delta_t\phi + \int_\Omega (\rho - \rho^*)\delta_t\rho$$

Applying the divergence theorem in the first term with the boundary condition $\nabla\cdot\boldsymbol{n} = 0$ we get

$$\dot{E} = -\int_\Omega |\nabla\cdot(\lambda\nabla\phi)|^2 - \int_\Omega \nabla\cdot(\lambda\nabla\phi)\kappa(\rho - \rho^*) + \int_\Omega (\rho - \rho^*)\delta_t\rho$$

Applying the continuity equation (3.5)

$$\dot{E} = -\int_\Omega |\nabla\cdot(\lambda\nabla\phi)|^2 - \int_\Omega \nabla\cdot(\lambda\nabla\phi)\kappa(\rho - \rho^*) - \int_\Omega (\rho - \rho^*)\nabla(\rho\boldsymbol{v})$$

And since $\boldsymbol{v} = -\kappa\frac{\lambda}{\rho}\nabla\phi$ we get

$$\dot{E} = -\int_\Omega |\nabla\cdot(\lambda\nabla\phi)|^2 - \int_\Omega \nabla\cdot(\lambda\nabla\phi)\kappa(\rho - \rho^*) + \int_\Omega (\rho - \rho^*)\kappa\nabla\cdot(\lambda\nabla\phi)$$

Then the last two terms cancel out and

$$\dot{E} = -\int_\Omega |\nabla\cdot(\lambda\nabla\phi)|^2$$

By Assumption 4 the orbits ϕ and ρ are bounded in $H^1(\Omega)$ and from Rellich-Kondrachov Theorem (2) the orbits are precompact in $L^2(\Omega)$. From La Salle Invariance Principle (3) the orbits converge to the largest invariant set in $E^{-1}(0)$. $\dot{E} = 0$ implies $||\nabla\cdot(\lambda\nabla\phi)||_{L^2(\Omega)} = 0$. From $\delta_t\phi = \nabla\cdot(\lambda\nabla\phi) + \kappa(\rho - \rho^*)$ we have that $\rho = \rho^*$ and the transport with vector $\boldsymbol{v} = -\kappa\frac{\lambda}{\rho}\nabla\phi$ converges $\rho \to \rho^*$ while ϕ reaches to optimality condition $\nabla\cdot(\lambda\nabla\phi) + \kappa(\rho - \rho^*) = 0$. □

With this theorem we can present the following algorithm

Algorithm 2 Multi-agent (on-the-fly) optimal transport with fixed (dual) weighting

Input: Target measure μ^*, transport cost $c(x, y)$, Weights (dual variable) λ_{ij}, Bound on step size ϵ, time step τ.

For each agent i at time instant k of transport:

1: Obtain positions $x_j(k)$ of neighbours within a communication radius r large enough to cover all Voronoi neighbours.

2: Compute Voronoi cell V_i, the mass of the Voronoi cell $\mu^*(V_i)$ and the Voronoi neighbour N_i

3: Initialize $\phi^i \leftarrow \Phi(k-1, x_i(k))$

4: Implement n iterations of the primal algorithm (5.15) synchronously in communication with $j \in N_i$ to obtain $\phi^i(k)$

5: Communicate with neighbors $j \in N_i$ to obtain $\phi^j(k)$ and construct a local estimate of $\Phi^d(k, x)$ with multivariate interpolation.

6: Implement transport step (5.3) with the local estimate $\Phi(k, x)$, which approximates $\phi_{\mu_k \to \mu^*}$

5.4 Simulations

We will now present the results of some simulations of Algorithm 2. We have implemented the algorithm in convex domains, as it was implemented in [30] and we have also provided new simulations in domains that present obstacles and are therefore not convex.
We start with the simulations in convex domains Ω. We have considered the Euclidean distance as a cost $c(x, y) = ||x - y||$, we have used $N = 30$ agents for the simulations and a value of $\epsilon = 0.05$. We have also considered a constant Lagrange multiplier $\lambda = \frac{1}{N_{max}} L$ with $N_{max} = 10$ and L the Laplacian of the graph $G = (\{x_i\}, E)$, with $(x_i, x_j) \in E$ if $\delta V_i \cap \delta V_j \neq \emptyset$. We have considered a target distribution with a covariance of $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ and mean randomly chosen in $[0, 1]^2$. The agents are initially distibuted normally following a gaussian law with a mean of $[5, 5]$ and covariance of $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$ In Figure 5.1 we can see how the agents move towards the target probability distribution and we converge to the target probability density.

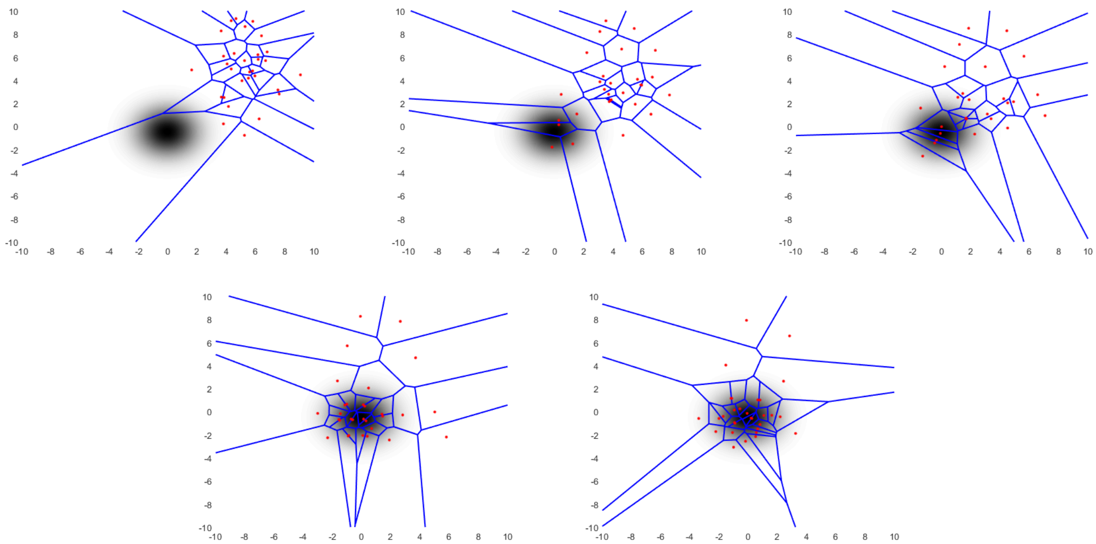

Figure 5.1: Evolution of a group of agents with their Voronoi partition towards the target probability distribution, shown in grayscale. We provide images at times $t = 0, 5, 10, 20, 30$

We can then study the convergence of the algorithm. We have implemented the primal-dual algorithm (2) with a time-step $\tau = 1$ using $n = 50$ iterations before convergence. We have measured the convergence using two metrics, the norm $||\rho - \rho^*||_{L^2(\Omega)}$ and the variance of the masses of the Voronoi cells $\mu^*(V_i)$. In Figure 5.2 we can see how the metrics decrease exponentially with time and how the constant κ can accelerate the convergence of the algorithm.

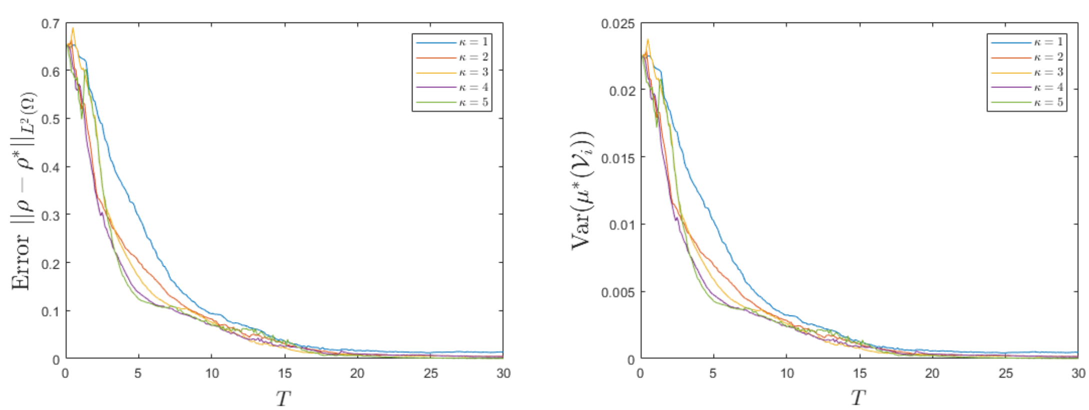

Figure 5.2: Convergence in the norm $||\rho - \rho^*||_{L^2(\Omega)}$ (left) and convergence with $Var(\mu^*(V_i))$ (right). We have plotted the results for $\kappa = 1, 2, 3, 4, 5$, we can see that the two norms decrease exponentially and we can increase the rate of convergence by increasing κ.

In our simulations we have observed that when $\kappa = 1$ the agents move towards the target probability measure following a random motion, when we increase κ the agents will find the desired direction easier, additionally we will converge to the target density measure with more precision, as we can see in Figure 5.3.

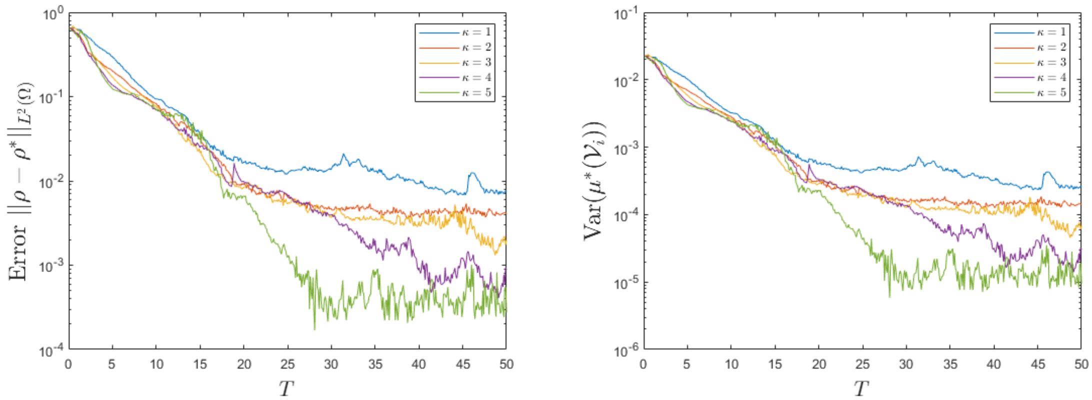

Figure 5.3: Convergence in the norm $||\rho - \rho^*||_{L^2(\Omega)}$ (left) and convergence with $Var(\mu^*(V_i))$ (right), both plotted in a logarithmic scale. We can see that higher values of κ will provide a more accurate convergence, with errors of the order of 10^{-3} for $\kappa = 4, 5$ while we achive a convergence of the order of 10^{-2} for $\kappa = 1, 2, 3$.

We will now focus on implementing the same algorithm when the domain Ω is not convex. We note that in non-convex domains the distance between points is given by the geodesic distance on the set,

$$d_g(x, y) = \inf_{\substack{\gamma(t):[0,1]\to\Omega \\ \gamma(0)=x,\, \gamma(1)=y}} \int_0^1 ||\dot{\gamma}(t)||dt$$

In order to implement the algorithm we would need to use a geodesic Voronoi partition, which can be done following [43]. We also note that in order to implement the algorithm with a geodesic Voronoi partition we would also need to estimate the geodesic distance in Ω and we would need a path planning algorithm to generate an obstacle free path from position $x_i(k)$ to position $x_i(k+1)$.
This is rather inconvenient and we have decided to work with an Euclidean Voronoi partition. From the Voronoi partition we can find the Voronoi graph $G_V = (\{x_i\}, E)$ with $(x_i, x_j) \in E$ if $\delta V_i \cap \delta V_j \neq \emptyset$. We will consider a subgraph of this graph, that only contains the edges that don't collide with obstacles, $\hat{G}_V = (\{x_i\}, \hat{E})$, where $(x_i, x_j) \in \hat{E}$ if $\delta V_i \cap \delta V_j \neq \emptyset$ and $z = (1-m)x_i + mx_j \in \Omega_{free}$ for $m \in [0, 1]$. We will call this graph the visibility Voronoi graph.
With the visibiility Voronoi graph we can define the constant Lagrange multiplier $\lambda = \frac{1}{N_{max}} L$ with $N_{max} = 10$ and L the Laplacian of the Visibility graph. Then we can use (5.8) to estimate the Kantorovich potential in a decentralized way. Finally we note that when we implement the transport scheme (5.3),

$$\begin{aligned} x_i(k+1) \in \arg\min_{z\in\Omega} \quad & ||x_i(k) - z|| + \phi_{\mu_k \to \mu^*}(z) \\ s.t \quad & ||x_i(k) - z|| \leq \epsilon \end{aligned}$$

we must ensure that $x_i(k+1) \in \Omega$ and that the points $(1-m)x_i(k) + mx_i(k+1) = z \in \Omega$ for all $m \in [0,1]$. As an optional improvement we can check if $\boldsymbol{v} = x_i(k+1) - x_i(k)$ doesn't move the agent towards an obstacles, we can check this by checking that $x_i(k) + \gamma \frac{\boldsymbol{v}}{||\boldsymbol{v}||} = z_i \in \Omega$ for all $\gamma \in [0,1]$.
Using this considerations we can run some simulations in an obstacle environment. We consider the same situation as in the previous experiment but we add a square centered in $[3.5, 3.5]$ with a heigh and witdh of $[1.5, 1.5]$. In Figure 5.4 we can see how the agents avoid the obstacle and converge to the desired location. We note that they require more time to converge since they have to avoid the obstacle. Finally, in Figure 5.5 we can see in a logarithmic plot how the error metrics decrease exponentially.

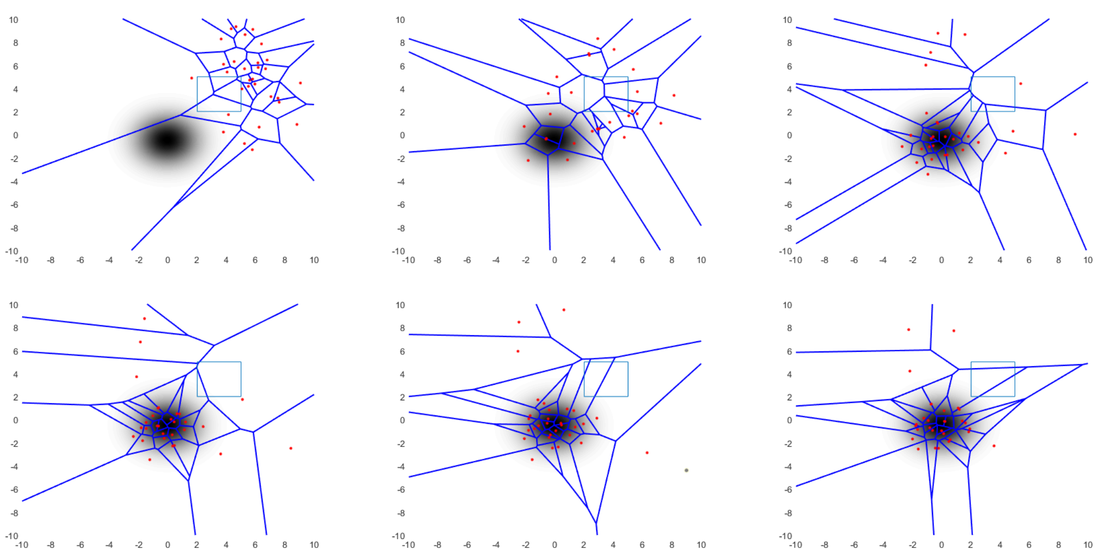

Figure 5.4: Evolution of a group of agents with their Voronoi partition towards the target probability distribution in an obstacle space, shown in grayscale. We provide images at times $t = 0, 10, 20, 30, 40, 50$

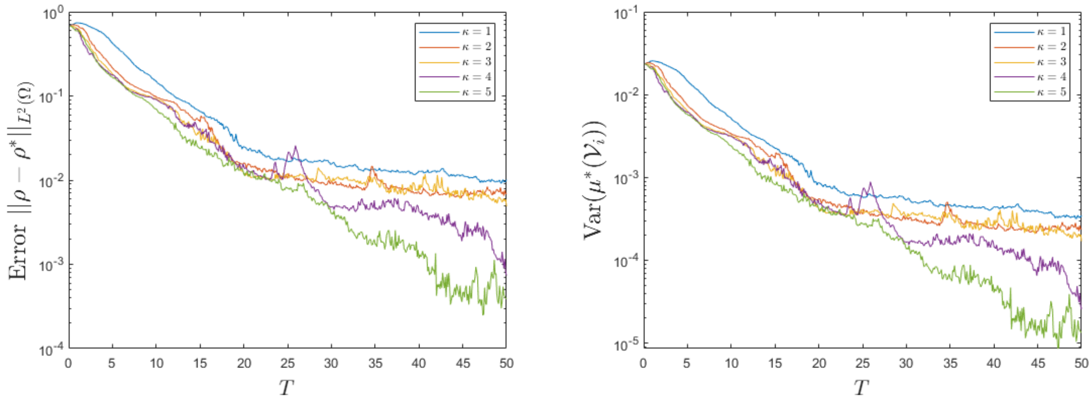

Figure 5.5: Logarithmic plot of the convergence in the norm $||\rho - \rho^*||_{L^2(\Omega)}$ (left) and convergence with $Var(\mu^*(V_i))$ (right) when considering an obstacle space. We note that we require $50s$ in order to achieve convergence, while we only required $30s$ to achieve convergence when we didn't consider the obstacle (5.3). Choosing a higher value of κ speeds up the convergence and gives us a more accurate solution.

5.5 Future Work: Collision Avoidance

We will now discuss some modifications that could be implemented in order to provide collision avoidance to the Multi-agent transport algorithm. We will propose three different models that could be adapted to our work. The models are unfinished and we leave their implementations as future work.
In the first model we consider interaction potentials that produce a repulsive force between the agents. In the second model we shortly talk about a Crowd Motion model that has been used in the literature to model microscopically the motion of agents. More research needs to be done in order to provide a good solution to the problem.

5.5.1 Interacting potentials

In the first model we will add a repulsive potential to the agents, following the ideas of [26]. We will consider a repulsive potential $W(r)$, which is decreasing with respect to r the distance between the agents. The potential of the agents defines an aggregate potential

$$V(x) = \sum_{i=1}^{n} W(x - x_i)$$

When we jump to the continuous by taking $n \to \infty$ we obtain the interaction potential

$$\mathcal{I}(\mu) = \int_{\Omega\times\Omega} W(x-y)d\mu(x)d\mu(y)$$

And our objective will be finding a gradient flow of the functional

$$W_2^2(\mu,\mu^*) + \mathcal{I}(\mu)$$

We now state some results about the interaction potential from [50].

Proposition 20. *If $W \in C_b(\Omega)$ then $\mathcal{I}$ is continuous for the weak convergence of probability measures. If W is lower semi-continuous and bounded from below then $\mathcal{I}$ is lower semi-continuous.*

We now analyze the displacement convexity of $\mathcal{I}$ in $\mathbb{W}_p$, where $\mu_t = (\pi_t)_{\#\gamma}$, with $\gamma(x,y)$ an optimal transport plan for the cost $c(x,y) = |x-y|^p$ and $\pi_t(x,y) = (1-t)x + ty$.

Proposition 21. *The functional $\mathcal{I}(\mu)$ is displacement convex if W is convex. Likewise, $\mathcal{I}(\mu)$ is displacement concave if W is concave.*

Proof.

$$\mathcal{I}(\mu_t) = \int W(x-x')(\pi_t)_{\#\gamma} d\gamma(x,y)(\pi_t)_{\#\gamma} d\gamma(x',y') =$$

$$= \int W((1-t)x + ty - (1-t)x' - ty')d\gamma(x,y)d\gamma(x',y')$$

$$= \int W((1-t)(x-x') + t(y-y'))d\gamma(x,y)d\gamma(x',y')$$

And if W is convex $W((1-t)(x-x') + t(y-y')) \leq (1-t)W(x-x') + tW(y-y')$, which give us

$$\leq \int (1-t)W(x-x') + tW(y-y')d\gamma(x,y)d\gamma(x',y') = (1-t)\mathcal{I}(\mu_0) + t\mathcal{I}(\mu_1)$$

Similarly if W is concave $W((1-t)(x-x') + t(y-y')) \geq (1-t)W(x-x') + tW(y-y')$ and we can prove

$$\mathcal{I}(\mu_t) \geq (1-t)\mathcal{I}(\mu_0) + t\mathcal{I}(\mu_1)$$

□

Since our objective is to minimize $W_2^2(\mu,\mu^*) + \mathcal{I}(\mu)$ we would like the functional to be convex. We note that the Wasserstein distance is convex in the generalized geodesics, as we have shown in (16) but the interacting functionals that we will consider will be concave. This may cause problems to achieve convergence. We leave for future work the analysis of the feasibility of using repulsive potentials.

5.5.2 Fixed collision radius

Other approaches have been proposed in the literature, specially in the field of Crowd Motion, that studies models of human motion in large gatherings. We remark the work of [37], and the adaptation of the work to the optimal transport formalism of [35], [36]. In their approach they treat the agents as rigid disks of radius r. They assume that each agent has a desired direction of movement v_i. This is enforced by considering a feasible initial configuration

$$Q = \{\boldsymbol{q} \in \mathbb{R}^{2n} : D_{ij}(\boldsymbol{q}) \geq 0 \quad \forall i, j\}$$

Where $\boldsymbol{q}$ is a vectorization of the positions of all agents and $D_{ij}(\boldsymbol{q}) = |q_i - q_j| - 2r$ corresponds to checking that agent i and agent j are separated by a greater distance than $2r$.
Then the feasible velocities are the ones that keep the initial position in the feasible set,

$$C_q = \{\boldsymbol{v} \in \mathbb{R}^{2n} : D_{ij}(\boldsymbol{q}) = 0 \implies G_{ij}(\boldsymbol{q}) \cdot \boldsymbol{v} \geq 0\}$$

With $G_{ij}(\boldsymbol{q}) = (0, \ldots, 0, -e_{ij}(\boldsymbol{q}, 0, \ldots, 0, -e_{ij}(\boldsymbol{q}, 0, \ldots, 0)$ and $e_{ij}(\boldsymbol{q}) = \frac{q_j - q_i}{|q_j - q_i|}$.
Meaning that when $|q_i - q_j| = 2r$ we will impose that the projection of $v_j - v_i$ in the direction e_{ij} is non-negative, and the agents don't get any closer.
With this they consider the transport given by

$$\dot{\boldsymbol{x}} = P_{C_q} \boldsymbol{v}$$

where P_{C_q} is a projection operator. This is traduced into they macroscopic transport

$$\delta_t \rho + \nabla \cdot (\rho P_{C_q} \boldsymbol{v}) = 0$$

We won't get into further details about this approaches but we are interested in analyzing how the algorithms perform for multi-agent robotic systems and also propose modifications.

5.5.3 Collision avoidance with bounds on the density function

We now present another approach that may be useful to consider and can be easily adapted to our work. We will consider continuous measures with a density, $\mu(x) = \rho(x)dx$, $\mu^*(x) = \rho^*(x)dx$ and we propose to add a bound on the maximum value of the density $\rho(x) \leq \rho_{max}$, with ρ_{max} a constant.
To implement this constraint we can define a Lagrangian according to

$$L(\mu, \lambda) = W_2^2(\mu, \mu^*) + F(\mu, \lambda)$$

with

$$F(\mu, \lambda) = \int \lambda(\rho^2 - \rho_{max}^2) dx$$

Here $\lambda > 0$ is the Lagrange multiplier, which is a function and we will assume $\lambda \in L^p(\Omega)$. We also note that the functional has the form

$$F(\mu, \lambda) = \int f_\lambda(\rho) dx$$

with $f_\lambda(\rho) = \lambda(\rho^2 - \rho_{max}^2)$.

Primal flow

We will now consider a primal-dual gradient flow on the Lagrangian $L(\mu, \lambda) = W_2^2(\mu, \mu^*) + F(\mu, \lambda)$, which depends on the density ρ and the function λ.
We start considering the minimization of $L(\mu, \lambda)$ over the primal variable ρ. We can formulate this minimization as a gradient flow in the Wasserstein space $\mathbb{W}_2$. We can take the following proximal gradient algorithm

$$\mu_{k+1} \in \arg\min_{\mu} W_2^2(\mu, \mu^*) + F(\mu, \lambda) + \frac{1}{2\tau} W_2^2(\mu, \mu_k)$$

We remember from the theory discussed in Chapter 3 that when the time-step $\tau \to 0$ the iterative scheme converges to the transport equation

$$\delta_t \rho_t + \nabla \cdot (\rho_t \boldsymbol{v_t}) = 0$$

with $\boldsymbol{v_t} = -\nabla \left(\frac{\delta L}{\delta \rho}(\rho, \lambda) \right)$. We will now calculate the first variation of $L(\mu, \nu)$.

$$\frac{\delta L}{\delta \rho}(\rho, \lambda) = \frac{\delta W_2^2}{\delta \rho}(\rho, \rho^*) + \frac{\delta F}{\delta \rho}(\rho, \lambda) = \phi_{\mu \to \mu^*} + f'_\lambda(\rho) = \phi_{\mu \to \mu^*} + 2\lambda\rho$$

Then the transport will be given by

$$\boldsymbol{v_t} = -\nabla \left(\phi_{\mu \to \mu^*} + 2\lambda\rho \right)$$

We will now remember from the theory that the transport equation is equivalent to the evolution of a group of particles, with initial positions distributed following the initial distribution $x_0 \sim \mu_0$ which evolve following

$$\begin{aligned} \dot{x}(t) &= -\nabla \phi_{\mu \to \mu^*}(x(t)) - 2\nabla(\lambda(x(t))\rho(x(t))) \\ x(0) &= x_0 \end{aligned} \tag{5.16}$$

In order to implement this gradient flow in discrete time we can follow a proximal gradient update

$$x_{k+1} \in \arg\min_{x} \phi_{\mu \to \mu^*}(x) + 2\lambda(x)\rho(x) + \frac{1}{2\tau} c(x, x_k) \tag{5.17}$$

Dual flow

We now present a the dual flow on λ. We note that $\lambda \in L^p(\Omega)$ is not a probability measure. In order to maximize the functional $L(\mu, \lambda)$ we will define a gradient flow, but since λ is not a probability measure the gradient flow won't be defined on the Wasserstein space, it will be defined in the $L^p(\Omega)$ function space. We start by defining the Fréchet derivative in L^p.

Definition 19 (Fréchet derivative in L^p)**.** *Let $\mathcal{F} : L^p(\Omega) \to \mathbb{R}$ be a functional we will call $\nabla_{L^p}\mathcal{F}$ the Fréchet derivative, if it exists, the functional $\nabla_{L^p}\mathcal{F} : L^p(\Omega) \to \mathbb{R}$ such that*

$$\lim_{\lambda \to \lambda_0} \frac{\mathcal{F}(\lambda) - \mathcal{F}(\lambda_0) - \nabla_{L^p}\mathcal{F}(\lambda - \lambda_0)}{||\lambda - \lambda_0||_{L^p}} = 0$$

We will denote the Fréchet derivative of the functional $L(\mu, \lambda)$ with respect to λ as $\nabla_\lambda L(\mu, \lambda)$. We can now write the dual gradient flow as

$$\delta_t \lambda = [\nabla_\lambda L(\mu, \lambda)]^+_\lambda$$

where $[f]^+_\lambda = \begin{cases} f & \text{if } \lambda > 0 \\ \max\{0, f\} & \text{if } \lambda = 0 \end{cases}$ is a projection operator. We can calculate the Fréchet derivative as

$$\nabla_\lambda L(\mu, \lambda) = \nabla_\lambda F(\mu, \lambda) = \rho^2 - \rho^2_{max}$$

And we can write

$$\delta_t \lambda = \left[\rho^2 - \rho^2_{max}\right]^+_\lambda \tag{5.18}$$

Future work

We leave for future work the analysis of the primal-dual flow that we have defined. We also note that in order to implement this algorithm will need to estimate the density function ρ and the Lagrange multiplier λ from the positions of the agents. Additionally the dynamics of λ must be discretized in space. A possibility could be to only study the dynamics of λ on the positions of the agents and interpolate the Lagrange multiplier on the other regions of space. This are early ideas and we have not had time to implement them or evaluate their feasibility.

Chapter 6

Conclusions and Future Work

In this thesis we have explored Coverage Control and Multi-Agent Transport of agents. We have studied the Time-Varying Coverage Control problem and a Singular Perturbation Theory approach to solve the Time-Varying Coverage Control. We have provided additional experimental validation to the algorithm and we have proposed a heuristic approach to ensure its convergence. We hope to publish our results soon.
We have presented the theory of Optimal Transport and we have related Optimal Transport with Coverage Control, finding that the Coverage Control locational cost is a relaxation of the Optimal Transport cost. Additionally we have see that Coverage Control cannot be studied with macroscopic models.
Then we have presented a recent approach to solve Multi-Agent Transport using a continuum model and Optimal Transport. We have contributed in the work increasing the convergence rate of the algorithms and implementing simulations when the domain presents obstacles.
We have also explored Collision Avoidance for the Multi-Agent Optimal Transport algorithm but our ideas still need refining. Our main approach of adding bounds to the density function is still incomplete, a proof of convergence for the primal-dual algorithm algorithm is needed, otherwise it might be interesting to consider other models. The discretization of the algorithm in space must also be studied carefully in order to provide a good implementation. Our other approaches to tackle Collision Avoidance are also promising, especially the fixed collision radius models that have been studied in the field of Crowd Motion. We believe that the Crowd motion algorithms could be adapted for Multi-Agent transport and the ideas presented in the Crowd Motion papers might lead to new approaches.

Bibliography

[1] Ambrosio, L. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces*. Springer, 2003, pp. 1–52.

[2] Ambrosio, L. Transport equation and cauchy problem for non-smooth vector fields. In *Calculus of variations and nonlinear partial differential equations*. Springer, 2008, pp. 1–41.

[3] Ambrosio, L., Gigli, N., and Savare, G. Gradient flows in metric spaces and in the space of probability measures, 2005.

[4] Ash, R. B., Robert, B., Doleans-Dade, C. A., and Catherine, A. *Probability and measure theory*. Academic Press, 2000.

[5] Bandyopadhyay, S., Chung, S.-J., and Hadaegh, F. Y. Inhomogeneous markov chain approach to probabilistic swarm guidance algorithm. In *5th Int. Conf. Spacecraft Formation Flying Missions and Technologies* (2013).

[6] Bandyopadhyay, S., Chung, S.-J., and Hadaegh, F. Y. Probabilistic swarm guidance using optimal transport. In *2014 IEEE Conference on Control Applications (CCA)* (2014), IEEE, pp. 498–505.

[7] Bandyopadhyay, S., Chung, S.-J., and Hadaegh, F. Y. Probabilistic and distributed control of a large-scale swarm of autonomous agents. *IEEE Transactions on Robotics 33*, 5 (2017), 1103–1123.

[8] Benamou, J.-D., and Brenier, Y. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik 84*, 3 (2000), 375–393.

[9] Billingsley, P. Convergence of probability measures, 2011.

[10] Bourne, D. P., and Roper, S. M. Centroidal power diagrams, lloyd's algorithm, and applications to optimal location problems. *SIAM Journal on Numerical Analysis 53*, 6 (2015), 2545–2569.

[11] Bullo, F., Cortés, J., and Martínez, S. *Distributed Control of Robotic Networks*. Applied Mathematics Series. Princeton University Press, 2009. Electronically available at http://coordinationbook.info.

[12] Cortés, J., Martínez, S., Karatas, T., and Bullo, F. Coverage control for mobile sensing networks: Variations on a theme. In *Mediterranean Conference on Control and Automation* (2002), Lisbon, Portugal Lisbon, Portugal, pp. 9–13.

[13] Cortés, J., Martínez, S., Karatas, T., and Bullo, F. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation 20*, 2 (2004), 243–255.

[14] de Badyn, M. H., Eren, U., Açikmeşe, B., and Mesbahi, M. Optimal mass transport and kernel density estimation for state-dependent networked dynamic systems. In *2018 IEEE Conference on Decision and Control (CDC)* (2018), IEEE, pp. 1225–1230.

[15] de Berg, M., Cheong, O., van Kreveld, M., and Overmars, M. *Computational Geometry: Algorithms and Applications*. Springer, 1997.

[16] Demir, N., Eren, U., and Açikmeşe, B. Decentralized probabilistic density control of autonomous swarms with safety constraints. *Autonomous Robots 39*, 4 (2015), 537–554.

[17] Dhariwal, A., Sukhatme, G. S., and Requicha, A. A. Bacterium-inspired robots for environmental monitoring. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004* (2004), vol. 2, IEEE, pp. 1436–1443.

[18] Diaz-Mercado, Y., Lee, S. G., and Egerstedt, M. Distributed dynamic density coverage for human-swarm interactions. In *2015 American Control Conference (ACC)* (2015), pp. 353–358.

[19] Drezner, Z. *Facility location: a survey of applications and methods*. Springer Verlag, 1995.

[20] Du, Q., and Emelianenko, M. Acceleration schemes for computing centroidal voronoi tessellations. *Numerical Linear Algebra with Applications 13*, 2-3 (2006), 173–192.

[21] Du, Q., Faber, V., and Gunzburger, M. Centroidal voronoi tessellations: Applications and algorithms. *SIAM Review 41*, 4 (1999), 637–676.

[22] Eren, U., and Açikmeşe, B. Velocity field generation for density control of swarms using heat equation and smoothing kernels. *IFAC-PapersOnLine 50*, 1 (2017), 9405–9411.

[23] Essid, M., and Solomon, J. Quadratically regularized optimal transport on graphs. *SIAM Journal on Scientific Computing 40*, 4 (Jan 2018), A1961–A1986.

[24] Galstyan, A., Hogg, T., and Lerman, K. Modeling and mathematical analysis of swarms of microscopic robots. In *Proceedings 2005 IEEE Swarm Intelligence Symposium, 2005. SIS 2005.* (2005), IEEE, pp. 201–208.

[25] Gandarillas, V. Singular perturbation method for multiagent coverage control using time-varying density functions. Master's thesis, UC San Diego, 2018.

[26] Howard, A., Matarić, M. J., and Sukhatme, G. S. Mobile sensor network deployment using potential fields: A distributed, scalable solution to the area coverage problem. In *Distributed Autonomous Robotic Systems 5*. Springer, 2002, pp. 299–308.

[27] Khalil, H. K. *Nonlinear Systems*. Prentice Hall, 1996.

[28] Krishnan, V. *Large-Scale Multi-Agent Transport: Theory, Algorithms and Analysis*. PhD thesis, UC San Diego, 2019.

[29] Krishnan, V., and Martínez, S. Distributed control for spatial self-organization of multi-agent swarms. *SIAM Journal on Control and Optimization 56*, 5 (2018), 3642–3667.

[30] Krishnan, V., and Martínez, S. Distributed online optimization for multi-agent optimal transport, 2018.

[31] Lee, S. G., Diaz-Mercado, Y., and Egerstedt, M. Multirobot control using time-varying density functions. *IEEE Transactions on Robotics 31*, 2 (2015), 489–493.

[32] Lévy, B. A numerical algorithm for l2 semi-discrete optimal transport in 3d. *ESAIM: Mathematical Modelling and Numerical Analysis 49*, 6 (2015), 1693–1715.

[33] Lloyd, S. Least squares quantization in pcm. *IEEE transactions on information theory 28*, 2 (1982), 129–137.

[34] Martinoli, A., Easton, K., and Agassounon, W. Modeling swarm robotic systems: A case study in collaborative distributed manipulation. *The International Journal of Robotics Research 23*, 4-5 (2004), 415–436.

[35] MAURY, B., ROUDNEFF-CHUPIN, A., and SANTAMBROGIO, F. A macroscopic crowd motion model of gradient flow type. *Mathematical Models and Methods in Applied Sciences 20*, 10 (Oct 2010), 1787–1821.

[36] Maury, B., Roudneff-Chupin, A., Santambrogio, F., and Venel, J. Handling congestion in crowd motion modeling. *Networks & Heterogeneous Media 6*, 3 (2011), 485–519.

[37] Maury, B., and Venel, J. A discrete contact model for crowd motion. *ESAIM: Mathematical Modelling and Numerical Analysis 45*, 1 (Jun 2010), 145–168.

[38] Mesbahi, M., and Egerstedt, M. *Graph theoretic methods in multiagent networks*. Princeton University Press, 2010.

[39] Mesquita, A. R., Hespanha, J. P., and Åström, K. Optimotaxis: A stochastic multi-agent optimization procedure with point measurements. In *International workshop on hybrid systems: Computation and control* (2008), Springer, pp. 358–371.

[40] Mérigot, Q. A multiscale approach to optimal transport. *Computer Graphics Forum 30*, 5 (2011), 1583–1592.

[41] Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, 2000.

[42] Okabe, A., and Suzuki, A. Locational optimization problems solved through voronoi diagrams. *European Journal of Operational Research 98*, 3 (1997), 445–456.

[43] Papadopoulou, E., and Lee, D.-T. A new approach for the geodesic voronoi diagram of points in a simple polygon and other restricted polygonal domains. *Algorithmica 20*, 4 (1998), 319–352.

[44] Pavone, M., Arsie, A., Frazzoli, E., and Bullo, F. Distributed algorithms for environment partitioning in mobile robotic networks. *IEEE Transactions on Automatic Control 56*, 8 (2011), 1834–1848.

[45] Peyré, G., and Cuturi, M. Computational optimal transport, 2018.

[46] Prorok, A., Correll, N., and Martinoli, A. Multi-level spatial modeling for stochastic distributed robotic systems. *International Journal of Robotics Research 30*, 5 (2011), 574–589.

[47] Rybski, P. E., Papanikolopoulos, N. P., Stoeter, S. A., Krantz, D. G., Yesin, K. B., Gini, M., Voyles, R., Hougen, D. F., Nelson, B., and Erickson, M. D. Enlisting rangers and scouts for reconnaissance and surveillance. *IEEE Robotics & Automation Magazine 7*, 4 (2000), 14–24.

[48] Santambrogio, F. Models and applications of optimal transport in economics, traffic and urban planning. *arXiv preprint arXiv:1009.3857* (2010).

[49] Santambrogio, F. Introduction to optimal transport theory. *Notes* (2014), 7.

[50] Santambrogio, F. Optimal transport for applied mathematicians, 2015.

[51] Santambrogio, F. Euclidean, metric, and Wasserstein gradient flows: an overview. *Bulletin of Mathematical Sciences 7*, 1 (Mar 2017), 87–154.

[52] Suzuki, A., and Drezner, Z. The p-center location problem in an area. *Location science 4*, 1-2 (1996), 69–82.

[53] Villani, C. *Optimal transport: old and new*, vol. 338. Springer Science & Business Media, 2008.

Appendix A

A counter example to the Twist Condition

Here we provide a counter example that shows that the Twist Condition (9) may not hold in non-convex domains. We start reminding the Twist condition (9),

Definition 20 (Twist condition)**.** *For $\Omega \subset \mathbb{R}^N$ we say $c : \Omega \times \Omega \to \mathbb{R}$ satisfies the Twist condition when c is differentiable with respect to x at every point and the map $y \to \nabla_x c(x_0, y))$ is injective for every x_0.*

When this condition holds we can deduce that if $(x_0, y_0) \in spt(\pi)$ then y_0 is uniquely defined and there exist an optimal transport map.

We will now show how this condition can fail when the domain is not convex. We consider the geodesic distance

$$d_g(x, y) = \inf_{\substack{\gamma(t):[0,1]\to\Omega \\ \gamma(0)=x,\, \gamma(1)=y}} \int_0^1 ||\dot{\gamma}(t)||dt$$

And we set $\Omega = [-10, 10] \times [-10, 10] \subset \mathbb{R}^2$, with an obstacle $O = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < 9\}$, which defines the compact free space $\Omega_{free} = \Omega \setminus O$. We can now consider a point x_0 and a point y_0 such that their linear interpolation $z = (1-m)x_0 + my_0$ for $m \in [0, 1]$ doesn't belong in Ω_{free} for all m. An example would be $x_0 = (-5, 1)$, $y_0 = (5, 1)$. It is easy to see that the geodesic γ^* between this two points connects x_0 with the boundary of O with a straight line that is tangential to the boundary of O. Then the geodesic follows the boundary of O until it connects with y_0 with a tangential straight line. In Figure A.1 we can this graphically.
We denote $z_1 = \gamma^*(t_1) \in \delta O$, and $z_2 = \gamma^*(t_2) \in \delta O$, where t_1, t_2 are the first and the last times such that the geodesic intersects with the boundary.

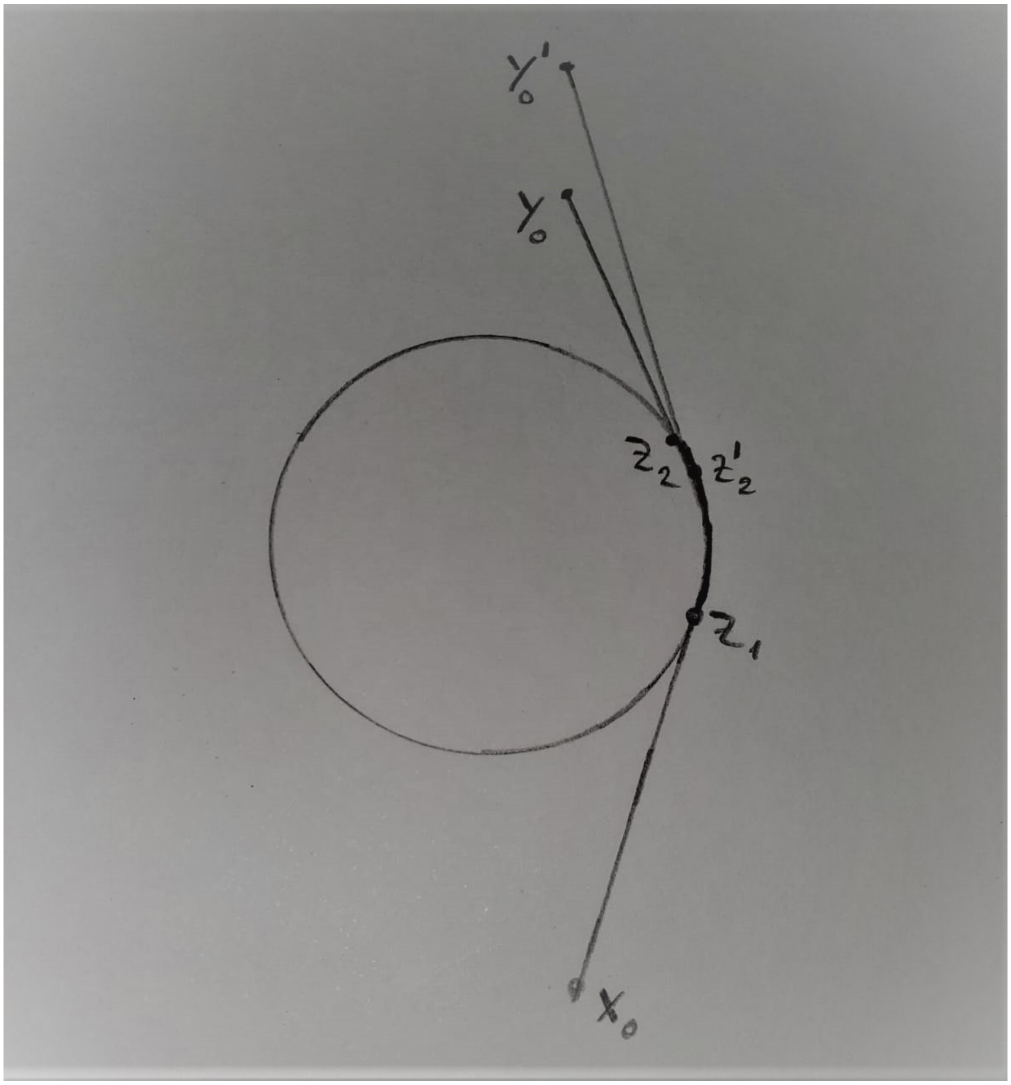

Figure A.1: Graphical representation of the setting

We then note that we can write the distance as

$$d_g(x_0, y_0) = ||x_0 - z_1|| + d_g(z_1, z_2) + ||z_2 - y_0||$$

Then,

$$\nabla_x d_g(x_0, y_0) = \nabla_x(||x_0 - z_1||) + \nabla_x(d_g(z_1, z_2))$$

And the gradient of the cost only depends on x_0 and z_1. Then we can choose $y_0' = (7, 1)$, which will lead to z_2' but x_0 and z_1 will stay the same and we will have

$$\nabla_x d_g(x_0, y_0) = \nabla_x d_g(x_0, y_0')$$

And the function $y \to \nabla_x d_g(x_0, y)$ is not injective, contradicting the Twist condition.