On metric theory of Gradient Flows

Applications in the Space of Probability Measures

Master Thesis

of

Georgios N. Domazakis

Advisor: Prof. Athanasios N. Yannacopoulos

School of Applied Mathematical and Physical Sciences
National Technical University of Athens
Greece

July 2019

Η μετρική θεωρία των Ροών Βαθμίδας

Εφαρμογές στον Χώρο των Μέτρων Πιθανότητας

Μεταπτυχιακή Εργασία

του

Γεωργίου Ν. Δομαζάκη

Επιβλέπων: Αθανάσιος Ν. Γιαννακόπουλος

Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών
Εθνικό Μετσόβιο Πολυτεχνείο
Ιούλιος 2019

Examination committee:

Athanasios. N. Yannacopoulos
Professor
Athens University of Econ. and Bus.
`ayannaco@aueb.gr`

Dimosthenis Drivaliaris
Assistant Professor
University of Aegean
`d.driv@fme.aegean.gr`

Nikolaos Yannakakis
Assosiate Professor
National Techical University of Athens
`nyian@math.ntua.gr`

To my family

To my mentor, A.N.Y

and

To whoose who care.

Contents

Foreword

Perhaps mathematics is not a science. This claim might be supported by the fact that there is no real world experiments in universe of mathematics. Possibly, mathematics is the *scientific language,* standing as a mental game, where anyone has to follow a certain class of rules, the *axioms,* where everything has to be synchronized with them. This synchronization of mathematical universe is governed by abstraction and generality. At the first glance, this abstraction might be useless and difficult to understand. Although, mathematics draw their strength from their abstract nature, and become able to treat situations, in a chic but also effective way, as a language. A language of logical thinking, whose systematic treatment was born more than two thousand years ago in the coasts of Aegean sea. So since mathematics is the language of logical and rational thinking, can the mathematical universe always overcome completely irrational situations? The answer of course is not. The history of such failures is pretty colourful. We can just name a few: Banach-Tarski paradox, Gödel's incompleteness theorem and Axiom of Choice may be the three most important examples of such a failure. Of course, we have to mention that these lines are pretend to by just a "hint" about mathematical universe and they are far from a complete epistemological approach.

However, the mathematical universe can provide a suitable framework to model several situation of real world. Everybody (or at least, *almost surely* everybody) is able to imagine and understand the impact of mathematics related to the progress of many scientific disciplines, like physics, biology, medicine, social sciences and so on. In many cases, such a framework presents a surprising and unexpected modelling success during last century and understanding the core components of a certain system. This success reminds us the famous treatise of Eugene Wigner back to 60s entitled as *"The unreasonable effectiveness of mathematics in natural sciences".*

In this direction, one of the greatest inventions (or possibly discoveries, depends on the philosophical school) inside the mathematical universe is the theory of *Differential Equations.* Their birthday essentially goes back to 17th century, under the hands of Isaac Newton, Gottfried Wilhelm Leibniz and Christian Huygens. Differential equations are associated with the *rate of change* of a certain quantity of a certain physical system. This physical system is associated with a specific mechanism, a model, which describes the quantities related with it. Let us describe the main idea of them by means of the following toy example. Imagine that we are a tiny leaf flying in the sky, due to air. We can drifting inside the air and flying and flying inside the air, taken by the flow of air. Wherever we are, we could calculate our rate of change of the quantities which describe our motion, like speed or acceleration, by solving the differential equation associated to our motion. This very simple analogy can be the starting point of understanding several physical phenomena, which involving rate of change.

It has been huge progress, which made us able to study, understand and master the physical laws, and create suitable models in order to understand complex phenomena. A very importnt part of this progress born the oldest siblings of Differential Equations, the so called *Partial Differential Equations.* These siblings allowed us to build more and more powerful models in order to study and explain more and more complex phenomena, like elasticity, gravitation, plasma physics, fluid mechanics and so on, and answer a class of many important questions, which partially might can justify the title of Wigner's paper. During last century, has been developed rigorous and solid mathematical theories, like Calculus of Variations and Nonlinear Analysis (see for example Brezis & Browder (1998) and Klainerman (2010) for a extended historical exposition).

Although, the theory of PDEs has not only provided suitable models to explain certain classes of systems. Essentially, they can provide us information backwards for the mathematical universe itself. Thought 20th century, the modern theory of PDEs was the primary and standard connection between the abstract mathematical universe and the physical world[1], where there was feedback to both sides. Let us mention two examples of such a connection. The first one can be described in terms of Differential Geometry. Precisely, questions related with minimal surfaces and embedding problems stands for a great motivation in order to study one of the most famous fully nonlinear PDEs, the *Monge-Ampére equation,* which stands until today an active research field. The second one stands on the bridge between PDEs and stochastics and probabilistic models. Essentially this connection was started by N. Wiener and the study Brownian motion, extended by Itô, Levy, Malliavin and many others to the theory of *Stochastic Differential Equations,* which gave arose to the stochastic counterpart of calculus of variations, the so-called Malliavin calculus. Nowadays, this connection is also a pretty active field research extended to non-local diffusions and Levy processes. The list of such connections is quite long and probably does not fit in the introductory lines. But, we have to mention that all of these connections has been useful and powerful both for physical models and abstract mathematical universe.

The aim of this thesis is to focus to a specific class of problems related with PDEs: the so-called *gradient flows.* Maybe the term "gradient flows" is originated to J. Hadamard and the *method of gradients* or *method of descent* for solving a PDEs by viewing them as a special case of equations with more variables tha initial one. Accurately, Hadamard mention (see Hadamard (1923)) the following:

> "*We thus have a first example of what I shall call a 'method of descent'. Creating a phrase for an idea which is merely childish and has been used since the first steps of the theory is, I must confess, rather ambitious; but we shall come across it rather frequently, so that it will be convenient to have a word to denote it. It consists in noticing that he who can do more can do less: if we can integrate equations with m variables, we can do the same for equations with $(m-1)$ variables.* "

Nowadays, Hadamard's method of gradients is inside the heard of the standard toolbox of any numerical analyst, employing it to solve numerically PDEs as heat equation and so on. Although, it was the starting point of a further extension of several problems related with PDEs. More precisely, it was the first way of understanding that behind of several problems there is a mechanism, the so-called *dissipation mechanism,* where the properties of which govern the corresponding system representing some sort of energy. Such a energy tends to minimizing, in order to drive the system towards its stable state. So, under this perspective, solving the PDE which is associated with the corresponding system is turned to minimization problem of a certain functional acting of certain spaces. A standard example of such a case is the heat equations, which can be recasted as the

[1]We avoid to use at this point the words *pure* and *applied* mathematics, since such a distinction in many situations does not exists in a concrete way!

minimization Dirichlet energy L^2. But, generally speaking, what about the functional and its acting spaces? During last decades, such problems was sightly studied. Accurately, during late 60s and 70s, it has been created beautiful theories by authors like H. Brezis, J.L. Lions, M.G. Crandall, T.M. Ligett and many others, treating the gradient flow problem in certain abstract vectorial settings, which can be putted under the title *Differential Inclusions and Maximal Monotone Operators in Hilbert spaces.*

Beyond vectorial setting, during last two decades the theory of gradient flows has been successfully developed in metric setting. Last decades, it has been developed a whole theory which make us able to recast a gradient flow problem in general *metric* setting, instead of the classical vectorial one. Such a procedure became possible for two specific reasons. The first one is related with the development of the Analysis in general metric spaces, where notions like *metric derivative, metric slope* and *geodesic convexity* has been developed, which turn the metric space setting into friendly one to work. The second one, is concentrated to the fact, that they were builded new functionals with desired properties, playing the role of dissipation mechanisms in certain metric space settings. Precisely, employing the fundamental principles of *Optimal Transport,* a beautiful mathematical theory which trace roots to G. Monge in 17th century and has been developed during 20th by L. Kantorivich, Y. Brenier, R. McCann and many others, we are able to define certain functional, with respect to which, we can recast many PDEs as gradient flows in the spaces of probability measures eqquiped with *Wasserstein distance.* The latter one has been sightly developed by the seminal work of R. Jordan, D. Kinderlehrer and F. Otto (see Jordan et al. (1998)). Moreover, the whole theory of gradient flows in space of probability measures is presented in the *the bible* of gradient flows by L. Ambrosio, N. Gigli and G. Savare (see L. Ambrosio et al. (2008)).

So, in this thesis, in order to present the metric theory of gradient flows we will follow the following path:

- In Chapter 1, we gently review the classical formulation of gradient flows in Euclidean setting. We present a numerical algorithm to solve such problems and finding three interesting and equivalent metric characterizations.
- In Chapter 2, we review some basic fact about Real Analysis and proceed to a further exploration to curves in metric spaces, recalling notions as *length spaces, geodesics in metric spaces* and *geodesic convexity,* which will be the first part of our standard toolbox in order to study gradient flows in metric space setting.
- In Chapter 3, we present at a glance, the optimal transport problem the general theory of Optimal Transportation, which will be the second part of our standard toolbox, in order to fight with gradient flows in metric space setting.
- In Chapter 4, we focus to the so-called *Wasserstein spaces,* which arose by optimal transport problem and enjoys several desirable properties.
- In Chapter 5, recalling the equivalent metric characterizations of gradient flows, we investigate the gradient flows in Wasserstein space for three classical functionals, employing elements of subddiferential calculus in Wasserstein space.

Acknowledgements

Trying to find the causality of this thesis or my enjoyable mathematical adventures of last years, all of my thoughts lead me to one single person: Professor Athanasios Yannacopoulos. I really feel lucky to met him a couple of years ago. Unluckily, during ten thousand years of mankind on Earth, there is no invented word which could express how grateful i feel to him, for his useful advises, his constant encouragement and his brilliant ideas. There is no doubt that this thesis could not be real without his existence (and of course, his uniqueness) as an unlimited source of inspiration. I sincerely thank him for all of our interesting discussions -mathematical or not- sometimes starting from philosophical issues, such as logical positivism in 30s or circle of Vienna and ending in a completely different direction with technical issues, such as linearised version Monge-Ampére equation or coarea formulas. I learn from him a lot! He gave me the opportunity to be a member of his research team, in AUEB, introducing me the general theory of Optimal Transportation, which has been the backbone of this thesis.

I am also thank Assistant Professor Dimosthenis Drivaliaris, for his continuously interesting of my postgraduate studies in National Technical University of Athens, especially about this thesis. He undercover to me a lot of chic functional analytic tools, in his way. Many thanks goes also to Associate Professor Nikolaos Yannakakis being in my examination board.

There are *three places* which played an important role in the development of this thesis. The first one is Mathematical Research Institute of Oberwolfach in Germany, where took place the Seminar *Optimal Transport Theory and Hydrodynamics (from Euler to Monge and vice versa)* in October 2018. I am glad to thank F. Santambrogio, M. Iacobelli and Y. Brenier whose give me the opportunity of participation and for their impressive lectures. The second one is Scuola Internazionale Superiore di Studi Avanzati (SISSA) in Trieste, where took place the winter school *Junior Math Days* in December 2018. The last one place, is Erwin Schrödinger International Institute for Mathematics and Physics in Vienna, where took place the *Thematic programme: Introductory School: Optimal Transport* in May 2019. Many thanks goes to M. Beiglböck, A. Figalli, J. Maas, R. McCann and J. Solomon for organizing such an interesting event and the opportunity they gave to participate. Of course, many thanks go to Lorenzo, Emanuelle, Veronica and Stefania for enjoyable moments in Oberwolfach, Trieste and Vienna.

There are also *three friends* who played a fundamental role in the progress of this thesis. First and foremost, i am glad to thank my great friend Dr. Georgios Papayiannis, where, before of couple of years, we were starting together an colourful investigation of the numerical aspects of Optimal Transport. Without any doubt, our discussions have played a crucial role to the construction of this thesis. I am glad also to thank Nicholas Protonotarios, who encouraning me to continue through the mathematical path, and for many enjoyable moments during postgraduate studies. Last but not least, I am glad to thank my brilliant officemate in, Katerina Mantzouni for spending plesant time together in Stochastic Modelling and Applications Laboratory.

In personal level, i thank my family, for their constant and unconditional support, both in moral and financial way.

Georgios N. Domazakis,
Athens,
July 2019

List of Symbols and Abbreviations

$B_r(x)$ (resp. $\overline{B_r(x)}$)	Open (resp. closed) ball with center $x \in \mathcal{X}$ and radius $r > 0$.
$\mathsf{p}_{\mathcal{X}}$	Canonical projection $\mathsf{p}_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathcal{X}$.
τ_d	Topology induced by the metric d in a metric space $(\mathcal{X}, d)$.
$\mathrm{Lip}(f)$	Minimal Lipschitz constant of a map $f : \mathcal{X} \to \mathcal{Y}$
$\mathrm{Lip}(\mathcal{X}, \mathcal{Y})$	Space of Lipschitz maps between two metric spaces $\mathcal{X}, \mathcal{Y}$.
$V_\gamma(a, b)$	Total variation of a curve $\gamma : I \to \mathcal{X}$, where I is an open interval.
$\ell(\gamma)$	Length of a curve $\gamma : I \to \mathcal{X}$, where I is an open interval.
$\mathscr{C}([a, b], \mathcal{X})$	Space of continuous paths $\gamma : [a, b] \to \mathcal{X}$ equipped with the topology of uniform convergence.
$L^p(\mathcal{X}, \mu)$	Lebesgue space on $\mathcal{X}$ with respect to the measure μ.
$AC(a, b; \mathcal{X})$ (resp. $AC_{loc}(a, b; \mathcal{X})$)	Space of absolutely continuous (resp. locally) curves on $[a, b]$ with values in $\mathcal{X}$.
$\lvert v' \rvert(t)$	Metric derivative of v at t
$D(\phi)$	Proper domain of a functional $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$.
$\mathscr{G}(X)$	Space of constant speed geodesics over the space $\mathcal{X}$ equipped with supremum norm.
e_t	Evaluation map$e_t : \mathscr{G}(\mathcal{X}) \to \mathcal{X}$ defined for any $t \in [0, 1]$.
$C^0_b(\mathcal{X})$	Space of continuous and bounded functionals defined on $\mathcal{X}$.
$C^\infty_c(\mathcal{X})$	Space of test function of a space $\mathcal{X}$.
$C^\infty_c(I \times \mathcal{X})$	Space of cylindrical test functions in space $\mathcal{X} \times I$, where $I \subset \mathbb{R}^d$, open with $d \geq 1$.
$W^{1,2}(\mathcal{D})$	Sobolev space of square integrable functions $f : \mathcal{D} \to \mathbb{R}$ where their weak derivative is square integrable.
$W^{1,2}_0(\mathcal{D})$	Closure of space of test functions $C^\infty_c(\mathcal{D})$ in $W^{1,2}(\mathcal{D})$
$\mathsf{Id}_{\mathcal{X}}$ or Id	Identical map of a metric space.$\mathcal{X}$
$\mathscr{M}(\mathcal{X})$	Space of (regular) Radon measures equipped with total variation norm.
$\mathcal{B}(\mathcal{X})$	σ-algebra of Borel subsets of the metric space. $\mathcal{X}$
$\mathscr{P}(\mathcal{X})$	Space of probability measures on $\mathcal{X}$.
$\mathscr{P}_p(\mathcal{X})$	Space of probability measures on $\mathcal{X}$ with finite p-moments.
$\boldsymbol{T}_\# \mu$	Push-foward measure of μ through the measurable map $\boldsymbol{T}$.
$\mathscr{L}^d$	$d-$dimensional Lebesgue measure.
$\mathscr{P}^{ac}_p(\mathcal{X})$	Space of probability measures in $\mathscr{P}_p(\mathcal{X})$ which is absolute continuous with respect to $\mathscr{L}^d$.
$\Pi(\mu, \nu)$	Set of transport plans between μ and ν.
$\mathcal{W}_p(\mu, \nu)$	$p-$Wasserstein distance between probability measures μ and ν.
$\mathbb{W}_p(\mathcal{X})$	$p-$Wasserstein space over a base space $\mathcal{X}$.
$\mathrm{Tan}_\mu \mathscr{P}_p(\mathcal{X})$	Tangent space of $\mathscr{P}_p(\mathcal{X})$ at point μ.

Abstract

It is the aim of this thesis to present the theory of gradient flows in metric setting and the space of probability measures. We revisit the classical vectorial theory, observing some powerful equivalent variational metric characterizations. Under the light of these, characterizations, we explore the theory of curves in metric spaces, which can be the suitable setting of such a generalization of vectorial environment Having this, we discuss, at a glance, the general theory of Optimal Transport, in order to end up with the so-called Wasserstein spaces, i.e. the space of probability measures equipped with Wasserstein distance, which enjoy several important metric, topological and geometrical properties. Under all of these considerations, we revisit the metric characterizations of gradient flows, where under suitable classes of assumptions, we are able to prove many desired results. Moreover, we conclude presenting three classical functionals, which produce gradient flows in Wasserstein space.

Περίληψη

Σκοπός της παρούσας διπλωματικής εργασίας είναι η παρουσίαση της μετρικής θεωρίας των ροών βαθμίδας και η εφαρμογές της στον χώρο μέτρων πιθανότητας. Εξετάζοντας προσεκτικά την διανυσματική θεωρία των ροών βαθμίδας, διαπιστώνουμε και περιγράφουμε διάφορους ενδιαφέροντες μεταβολικούς χαρακτηρισμούς, οι οποιοι οτι μας επιτρέπουν μια μετρική γενίκευση του διανυσματικοί πλαισίου. Ακολουθώντας αυτήν την παρατήρηση, μελετάμε τις καμπύλες σε μετρικούς χώρους, βασιζόμενοι σε αρκετά εργαλεία απο την Μετρική Γεωμετρία και την Ανάλυση σε Μετρικούς χώρους. Έχοντας επικαλεστεί τις απαραίτητες έννοιες, κάνουμε μια σύντομη παρουσίασή της θεωρίας της Βέλτιστης Μεταφοράς, και τους χώρους και τις αποστάσεις Wasserstein, οι οποίοι εμφανίζουν πληθώρα σημαντικών μετρικών, τοπολογικών και γεωμετρικών ιδιοτήτων. Μέσα από το πρίσμα όλων αυτών, παρουσιάζουμε την μετρική θεωρία των ροών βαθμίδας, όπου κάτω απο συγκεκριμένες κλάσεις υποθέσεων είμαστε σε θέση να διατυπώσουμε αρκετά επιθυμητά αποτελέσματα. Τέλος, παρουσιάζουμε 3 κλασσικά συναρτησιακά στον χωρο Wasserstein, παράγοντας συγκεκριμένες ροές βαθμίδας.

CHAPTER 1

Introduction

In this introductory chapter, we will revisit the notion of gradient flows in vectorial setting, presenting the gradient flow equation in $\mathbb{R}^d$. We are going to one of archetypes of evolution systems: *the heat equation* as a gradient flow in L^2. Up to these definitions, we will present a class of useful variational characterizations of gradient flow equation, as well as, its approximation by a time-discretization procedure and its equivalent metric characterizations, which will be our stepping stone for our further exploration.

1.1 Motivation: evolution through time

Perhaps, one could say that the real world is full-filled of optimization problems. Very often, people in order to take a decision, to minimize some cost or time, or even to maximize some profit, need to optimize some quantity which is related to their problem. The same fact holds also for nature. The physical world is also filled with optimization problems. Physical systems tends to a state of minimum energy with respect to a specific mechanism, a physical principle, the so called *dissipation mechanism.* Roughly speaking, masses "want" to minimize their potential energy, rays of light follow paths that minimize their travel time, molecules in a isolated chemical system react with each other until the total potential energy of their electrons is minimized, and so on. Although, what is the common characteristic of all above optimization problems? *Their evolution in time.*

So, attracting a wide field of interest, optimization problems are foremost in modern science. They could be divided in several classes, depending on their mathematical treatment and enjoying many interesting approaches. Of course, these lines are not too wide to be filled with a generic exposition of the modern ways of treatment optimization problems. However, a very interesting fact about physical systems, where some quantity has to be optimized, is the study their time evolution. In particular, under this perspective the study of such systems in concentrated to the description of the evolution of their characteristic quantities through time, which conceptually described via the dissipation mechanism.

In several situations, such an evolution can be translated in terms of Partial Differential Equations (PDEs), having as a stepping stone, the dissipation mechanism which is connected with the corresponding physical phenomenon. This translation provide to us several hints to understand the underlying dynamical structure (or *the gradient structure*) of that physical phenomenon and many other important qualitative and quantitative properties, due to generality and flexibility of this PDE approach.

Let us proceed, by making the above discussion a little bit more concrete. For that reason, restricting ourselves to the Euclidean setting, we consider a smooth enough functional $\phi : \mathbb{R}^d \to \mathbb{R}$ and the minimization problem.

$$\min_{x \in \mathbb{R}^d} \phi(x). \tag{1.1}$$

To describe its infinitesimal change of rate of ϕ, we recall the classical the notion of *gradient* of ϕ, which is defined as $\nabla\phi := \left(\frac{\partial\phi}{\partial x_1}, \cdots, \frac{\partial\phi}{\partial x_d}\right)^\top$. In the spirit of gradient of ϕ, we recall its connection with directional derivative, i.e.

$$\frac{d}{dt}\phi(x_0 + th) = \langle h\nabla\phi(x)\rangle, \qquad \text{for } h \in \mathbb{R}^d$$

based on which, we can derive the following ODE system

$$\frac{d}{dt}x(t) = -\nabla\phi(x(t)). \tag{1.2}$$

Upon a suitable differentiability assumption of (1.2) has the property

$$\frac{d}{dt}\phi(x(t)) = -\langle\nabla\phi, \phi\rangle \leq 0,$$

which means that it is decreasing along the orbits of the ODE (1.2). Therefore, starting at the point x_0 at $t_0 = 0$, the orbit $x(t; t_0)$ moves along a path (or a curve) such that as t increases $\phi(x(t))$ decreases. Under a supplementary condition, one may convenient that this orbit asymptotically leads to a minimum of the functional ϕ. This observation has led us to a number of interesting numerical approach to the minimization problem (1.1) having as their basic concept the evolution of the ODE system (1.2), often called *gradient flow* (or gradient system), in order to provide a subsequent approximation of the functional under consideration. The above convenient gradient structure can be extended in a infinite dimensional separable Hilbert space setting (see e.g. Brézis (1971)).

However, having the above heuristic description of gradient flows in our mind, we are considering about one basic examples of such a gradient flow structure. One of the backbone examples one may consider, reveals one of the four fundamental equations of mathematical physics, *the heat equation,* and its connections with the minimization of Dirichlet energy functional. This equation can be viewed as a dissipative evolution equation, or accurately, as the result of the principle of conservation of total heat and Fourier's law.

The heat equation has the typical gradient structure that was described in above discussion. To see this, let us the forthcoming discussion a little bit more technical. For that reason, let us assume that $\mathcal{D} \subset \mathbb{R}^d$ be be an smooth domain (i.e. open and connected) and the following boundary value problem for the heat equation:

$$\begin{aligned} \frac{d}{dt}u - \Delta u &= 0, \qquad \text{in } \mathcal{D} \times (0, \infty) \\ u(0) &= u_0, \qquad \text{in } \mathcal{D} \end{aligned}$$

where $u_0 \in \mathbb{R}^d$ stands for the initial condition.

Let us recall now the basic functional analytic machinery to treat the above problem. We recall that $W^{1,2}(\mathcal{D})$ denotes the Sobolev space of squared integrable functions with respect to Lebesgue measure such that their second weak derivative exists, that is

$$W^{1,2}(\mathcal{D}) := \left\{u \in L^2(\mathcal{D}) : Du \in L^2(\mathcal{D})\right\},$$

equipped with the norm $\|v\|_{W^{1,2}(\mathcal{D})} = \Big(\|u\|^2_{L^2(\mathcal{D})} + \|Du\|^2_{L^2(\mathcal{D})}\Big)^{1/2}$. We consider also the subspace $W^{1,2}_0(\mathcal{D}) \subseteq W^{1,2}(\mathcal{D})$ which is defined as the closure of space of test functions $C^\infty_c(\mathcal{D})$ in $W^{1,2}(\mathcal{D})$, or accurately, we will have that $u \in W^{1,2}_0(\mathcal{D})$ if and only if there exists a sequence $(u_n)_{n\in\mathbb{N}} \subset C^\infty_c(\mathcal{D})$ such that

$$\|u_n - u\|_{W^{1,2}(\mathcal{D})} \to 0, \quad \text{as} \quad n \to \infty.$$

After the functional analytic "refuelling stop" let us return to the problem of the gradient flow structure of heat equaton. We consider *the Dirichlet energy functional* $\mathcal{E} : W^{1,2}_0(\mathcal{D}) \to \mathbb{R} \cup \{+\infty\}$ defined as

$$\mathcal{E}(u) = \int_{\mathcal{D}} |\nabla u|^2 dx. \tag{1.3}$$

The Gateaux derivative of $\mathcal{E}$ is defined as the operator

$$\tag{1.4}$$

$$d\mathcal{E}(u)[f] = \lim_{t\to 0} \frac{\mathcal{E}(u + t\ f) - \mathcal{E}(u)}{t} \tag{1.5}$$

which is well-defined when f belongs to $L^2(\mathcal{D})$. Equivalently, we ask some $f \in W^{1,2}_0(\mathcal{D})$, such (1.5) is reduced to

$$\mathcal{E}(u + tf) = \mathcal{E}(u) + t\int_{\mathcal{D}} \nabla u \cdot \nabla f dx + t^2\mathcal{E}(f). \tag{1.6}$$

So we can understand that the functional $\mathcal{E}'(u)[\cdot] : W^{1,2}_0(\mathcal{D}) \to \mathbb{R}$ defined as

$$d\mathcal{E}(u)[f] := \int_{\mathcal{D}} \nabla u \cdot \nabla f dx, \quad \text{for } f \in W^{1,2}_0(\mathcal{D}) \tag{1.7}$$

consists a linear operator. In the same fashion, we can define the gradient of $\mathcal{E}$ at u, representing $dE(u)$ via a duality pairing. Thanks to Riesz representation theorem, there exists a unique $\nabla_{L^2}\mathcal{E}(u) \in L^2(\mathcal{D})$ such that

$$\nabla_{L^2}\mathcal{E}(u) = \langle d\mathcal{E}(u)[f], f\rangle_{L^2}, \quad \text{for all } f \in L^2(\Omega). \tag{1.8}$$

Employing integration by parts, we can easily obtain

$$d\mathcal{E}(u)[f] = \int_\Omega \nabla u \nabla f dx = -\int_\Omega f\Delta u dx \tag{1.9}$$

and based in this fact we conclude that $\nabla_{L^2}\mathcal{E}(u) = -\Delta u$. Under this perspective, we can reduce to the analogue of the ODE (1.2) for the heat equation, as gradient flow in L^2 in the following form

$$\frac{d}{dt}u(t) = -\nabla_{L^2}\mathcal{E}(u(t)). \tag{1.10}$$

So, in the light of above routine calculations, we realize that the heat equation consists a flow in a specific spaces of functions. We will call the above equation as *gradient flow* (or gradient system) in Hilbert space setting, and in particular in L^2 setting, since the inner product defining the Gateaux derivative is used. Moreover, note that (1.10) can be considered as a generated flow in L^2, as it can by operator semigroup theory. In this spirit, using semigroup theory, we will be able to prove several results about gradient structures in what follows.

1.2 Existence and fundamental estimates

Since now, our discussion related to gradient flows was in a quite intuitive and heuristic manner. As we have already seen, one of the fundamental PDEs, the heat equation, can be expressed as a minimization problem in a specific function space and a specific dissipation mechanism, the Dirichlet energy, and hence, functional analytic methods could be employed to study several important properties of them.

We shall turn our discussion to a more rigorous one, in order to expose some important results related to uniqueness of gradient flows and variational characterizations of them, as well as their fundamental estimates. For simplicity, we will restrict ourselves to Euclidean setting. Note that all of the above results could be stated and proved in any separable (finite or infinite) Hilbert space.

We will start this exploration by recalling the definition of λ-*convexity,* a technical relaxation of classical notion of convexity in $\mathbb{R}^d$ which makes us able to recast nice equivalent variational characterizations.

Definition 1.2.1 (λ-convexity)**.** A functional $\phi : \mathbb{R}^d \to \mathbb{R}$ is called λ-convex (or semiconvex) if the map $x \mapsto \phi(x) - \frac{\lambda}{2}|x|^2$ is convex for some $\lambda \in \mathbb{R}$.

The intuition behind the above relaxation of classical convexity is concentrated to two facts. On the one hand, one can observe that for $\lambda > 0$, the concept of λ-convexity is stronger than usual convexity. Indeed, imagine that based on Definition 1.2.1, λ-convex functionals for $\lambda > 0$, remain convex, even if we subtract a positive multiplier in the form of the function $x \mapsto \frac{|x|^2}{2}$. On the other hand, λ-convex functionals for $\lambda < 0$, actually need to be added with a positive multiplier in the form of $x \mapsto \frac{|x|^2}{2}$ to become convex, so in this case λ-convexity is weaker that usual convexity. This technical relaxation of convexity turn out to be very useful for treating various situation of our interest, as we will see in the following.

In the following proposition, we present four extremely useful equivalent characterizations of $\lambda-$convexity, under some smoothness assumption, which will play crucial role what follows this section, providing some fundamental estimates.

Proposition 1.2.2 (Equivalent characterizations of λ-convexity)**.** *Consider that $\phi \in C^2(\mathbb{R}^d)$ is a λ-convex functional, $x, x_0, x_1 \in \mathbb{R}^d$ and where x_w is defined as the convex combination of x_0 and x_1, that is, $x_w := (1-w)x_0 + wx_1$, for some $w \in [0,1]$. Then the following are equivalent:*

(i) *λ-monotonicity of $\nabla\phi$:*

$$\langle \nabla\phi(x_0) - \nabla\phi(x_1), \phi(x_0) - \phi(x_1) \rangle \geq \lambda |x_0 - x_1|^2. \tag{1.11}$$

(ii) *λ-convexity inequality :*

$$\phi(x_w) \leq (1-w)\phi(x_0) + w\phi(x_1) - \frac{\lambda}{2} w(1-w)|x_0 - x_1|^2 \tag{1.12}$$

(iii) *Hessian inequality:*

$$D^2\phi(x) \geq \lambda \mathsf{Id} \tag{1.13}$$

(iv) *Subgradient inequality:*

$$\langle \nabla\phi(x_1), x_1 - x_0 \rangle - \frac{\lambda}{2}|x_1 - x_0|^2 \geq \phi(x_1) - \phi(x_2) \geq \langle \nabla\phi(x_0), x_1 - x_0 \rangle + \frac{\lambda}{2}|x_1 - x_0|^2. \tag{1.14}$$

The proof of this proposition follows by a typical application of the definition of λ-convexity. For that reason it omitted. Nevertheless, shall we make some comments on theses characterizations. First of all, the λ-monotonicity of $\nabla\phi$ essentially describes nothing more but a technical relation of the concept of *monotone operator.* In addition, the λ-convexity inequality and Hessian inequality describe how this technical relaxation works, by means of a correction, in some sense, of a convex function. Finally, the subgradient inequality, witnesses the interplay between λ-convexity and the differential structure of $\mathbb{R}^d$.

Having already a relaxation of usual convexity in our toolbox, let us define more formally the notion of gradient flow in $\mathbb{R}^d$.

Definition 1.2.3 (Gradient flow in $\mathbb{R}^d$)**.** Consider that $\phi : \mathbb{R}^d \to \mathbb{R}$ is a smooth enough functional. The gradient flow of a map ϕ is the family of maps

$$\mathcal{S}_t : \mathbb{R}^d \to \mathbb{R}^d, \quad \text{for } t \in [0, \infty),$$

where for every $u_0 \in \mathbb{R}^d$ with $\mathcal{S}_0(u_0) := u_0$ and the map $u_t := \mathcal{S}_t(u_0)$ is the unique C^1-solution to the Cauchy problem

$$\frac{d}{dt}u_t = -\nabla\phi(u_t) \quad \text{in } (0,\infty) \text{ with } u_t \to u_0, \text{ as } t \downarrow 0. \tag{1.15}$$

It is widely known that, one can employ the Cauchy-Lipschitz theory and *a priori* estimates, in order to show that for every initial datum $u_0 \in \mathbb{R}^d$ the equation (1.15), admits a unique (global) solution. Moreover, the family of maps $(\mathcal{S}_t)_{t\geq 0}$ defining a continuous semigroup of Lipschitz maps, i.e. satisfying the properties

$$\mathcal{S}_{t+h}(u_0) = \mathcal{S}_t(\mathcal{S}_h(u_0)) \quad \text{and} \quad \mathcal{S}_t(u_0) = \mathcal{S}_0(u_0) = u_0, \text{ as } t \downarrow 0 \text{ and for every } u_0 \in \mathbb{R}^d. \tag{1.16}$$

So, based on above definition and using some smoothness and the λ-convexity assumption on functional $\phi : \mathbb{R}^d \to \mathbb{R}$, we can derive three very important estimates for gradient flows, which will play a vital role in what follows. Let us present them in the following proposition.

Proposition 1.2.4 (Fundamental estimates)**.** *Consider that the functional $\phi : \mathbb{R}^d \to \mathbb{R}$ is C^2 and λ-convex. Let also $u : [0,\infty) \to \mathbb{R}^d$ is a solution of* (1.15)*. Then the following hold:*

(i) Evolution Variational Inequality:

$$\frac{d}{dt}\frac{1}{2}|u_t - v|^2 + \frac{\lambda}{2}|u_t - v|^2 = e^{-\lambda t}\frac{d}{dt}\big(e^{\lambda t}\frac{1}{2}|u_t - v|^2\big) \leq \phi(v) - \phi(u_t) \quad \textit{for every } v \in \mathbb{R}^d \tag{1.17}$$

(ii) Energy Identity:

$$\frac{d}{dt}\phi(u_t) = -\Big|\frac{d}{dt}u_t\Big|^2 = -|\nabla\phi(u_t)|^2 \leq 0 \tag{1.18}$$

(iii) Slope inequality:

$$\frac{d}{dt}\Big(e^{2\lambda t}|\nabla\phi(u_t)|^2\Big) = \frac{d}{dt}\Big(e^{2\lambda t}\Big|\frac{d}{dt}u_t\Big|^2\Big) \leq 0 \tag{1.19}$$

Proof. The proof of all of these fundamental estimates rely on simple arguments based on the equivalent characterizations of a $\lambda-$convex functional, provided by Proposition 1.2.2.

(i) For the evolution variational inequality, by the definition of gradient flow in (1.15) and subgradient inequality given by 1.14, we have

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}|u_t - v|^2 &= \langle \frac{d}{dt}u_t, u_t - v\rangle \\ &\overset{(1.15)}{=} \langle -\nabla\phi(u_t), u_t - v\rangle \\ &= \langle \nabla\phi(u_t), v - u_t\rangle \\ &\overset{(1.14)}{\leq} \phi(v) - \phi(u_t) - \frac{\lambda}{2}|u_t - v|^2 \quad \text{for every } v \in \mathbb{R}^d.\end{aligned}$$

(ii) For the energy identity, again by the definition of gradient flow in (1.15), we have

$$\frac{d}{dt}\phi(u_t) = \left\langle \nabla\phi(u_t), \frac{d}{dt}u_t \right\rangle \overset{(1.15)}{=} -|\nabla\phi(u_t)|^2 \leq 0.$$

(iii) For the slope inequality, we employ again the definition of gradient flow in (1.15) and the Hessian inequality given by (1.13), we have

$$\begin{aligned}\frac{d}{dt}|\nabla\phi(u_t)|^2 &= 2\left\langle D^2\phi(u_t)\nabla\phi(u_t), \frac{d}{dt}u_t \right\rangle \\ &\overset{(1.15)}{=} -2\left\langle D^2\phi(u_t)\nabla\phi(u_t), \nabla\phi(u_t) \right\rangle \\ &\overset{(1.13)}{\leq} -2\lambda|\nabla\phi(u_t)|^2\end{aligned}$$

and based on this, we can obtain the desired bound. □

Remark 1.2.5 (λ-contraction property)**.** Let us remark now an important consequence of the Proposition 1.2.4, where due to its assumptions, we can obtain a λ-contraction property. Precisely, if $v \in \mathbb{R}^d$ is another solution to the gradient flow problem (1.15), then it holds that

$$\frac{d}{dt}\left(e^{\lambda t}|u_t - v_t|\right) \leq 0. \tag{1.20}$$

To see this, if we employ again the definition of gradient flow of (1.15) and λ-monotonicity given by (1.11), we can derive

$$\begin{aligned}\frac{d}{dt}|u_t - v_t|^2 &= 2\left\langle \frac{d}{dt}u_t - \frac{d}{dt}v_t, u_t - v_t \right\rangle \\ &\overset{(1.15)}{=} -2\langle \nabla\phi(u_t) - \nabla\phi(v_t), u_t - v_t\rangle \\ &\overset{(1.11)}{\leq} -2\lambda|u_t - v_t|^2,\end{aligned}$$

and based on this, the desired bound follows.

Following the spirit of **?**, we can recast useful and equivalent integral versions of the fundamental estimates. These versions could be applied to prove existence, uniqueness and stability of the flow under our assumptions. We highlight this fact in the following remark.

Remark 1.2.6 (Fundamental estimates-revised and well-poseness)**.** The three basic estimates of Proposition 1.2.4 can be translated in integral form which will be useful in order to understand better the Lipschitz property of $\mathcal{S}_t$. Let us make this claim more rigorous by setting

$$G_\lambda(t) := \int_0^t e^{\lambda r}dr = \begin{cases} \frac{e^{\lambda t}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ t, & \text{if } \lambda = 0 \end{cases}. \tag{1.21}$$

Then, based on (1.21), the *evolution variational inequality* becomes

$$\frac{e^{\lambda t}}{2}|u_t - u|^2 + G_\lambda(t)\big(\phi(u_t) - \phi(v)\big) + \frac{\big(G_\lambda(t)\big)^2}{2}|\nabla\phi(u_t)|^2 \le \frac{1}{2}|u_0 - v|^2, \quad \text{for every } v \in \mathbb{R}^d,$$

the *energy identity* becomes

$$\phi(u_t) + \frac{1}{2}\int_0^t \Big(\Big|\frac{d}{dr}u_r\Big|^2 + |\nabla\phi(u_r)|^2\Big)dr = \phi(u_0),$$

and *slope inequality* becomes

$$|\nabla\phi(u_t) \le e^{-\lambda t}|\nabla\phi(u_0)|.$$

Moreover, after a short period of calculations, we can observe that if $v \in \mathbb{R}^d$ is another solution to (1.15), then we can derive

$$|u_t - v_t| \le e^{-\lambda t}|u_0 - v_0|. \tag{1.22}$$

In addition, translating in (1.22) in terms of maps $\mathcal{S}_t$ in definition of gradient flow, we obtain an λ-contraction-type estimate in the form

$$d(\mathcal{S}_t(u_0), \mathcal{S}_t(v_0)) \le e^{-\lambda t}d(u_0, v_0), \quad \text{for every} u_0, v_0 \in \mathbb{R}^d, \quad t \in [0, \infty).$$

which shows the Lipschitz property of $\mathcal{S}_t$ and thus uniqueness and stability of solution of (1.15).

1.3 Approximation of solutions by time-discretization

Since now, we have explored the well-poseness of gradient flow equation, which was a standard application of the theory semigroups of linear operators, based on our smoothness and convexity assumptions. A natural next step of this exploration is to find a suitable way to approximate solutions of gradient flow equation. Perhaps, one of the simplest, but very useful, ways to produce such approximations of this type is the so-called *implicit Euler scheme.* According to it, we can construct discrete approximations of the solutions of gradient flow equations, using a time-discretization, and show its existence by a limiting process.

Let us make this discussion a little bit more rigorous. Consider a time step parameter $\tau > 0$ and the associated uniform partition of $[0, +\infty)$ in the form

$$\mathcal{P}_\tau := \{0 = t_\tau^0 < t_\tau^1 < \cdots < t_\tau^n < \cdots\}, \quad \text{where } t_\tau^n := n\tau. \tag{1.23}$$

One goal is to find a discrete sequence $(U_\tau^n)_{n\in\mathbb{N}}$, whose value U_τ^n has to provide an effective approximation $u(t_\tau^n)$. Under this consideration, we can define the term U_τ^n recursively, starting from a suitable choice of $U_\tau^0 \approx u_0$, by solving at each step the well-known implicit Euler scheme

$$\frac{U_\tau^n - U_\tau^{n-1}}{\tau} = -\nabla\phi(U_\tau^n), \quad n = 1, 2, \cdots, \tag{1.24}$$

or equivalently,

$$U_\tau^n = J_\tau(U_\tau^{n-1}), \tag{1.25}$$

where $J_\tau := (\mathsf{Id} + \tau\nabla\phi)^{-1}$ is the resolvent operator for the gradient flow.

Under the light of this approximation procedure, we can obtain the existence of a discrete approximating solution by looking for minimizers of the functional

$$U \longmapsto \phi(U) + \frac{1}{2\tau}|U - U_\tau^{n-1}|^2 =: \Phi(\tau, U_\tau^{n-1}, U_\tau^n). \tag{1.26}$$

To see this, we have only to check that any minimizer U_τ^n of (1.26) solves the equation (1.24), since the functional defined in (1.26) is $(\tau^{-1}+\lambda)-$convex and therefore it admits a unique minimizer, as long as $\tau^{-1} > -\lambda$, by a standard application of Weierstrass theorem.

But what about the convergence of this scheme? Could we have any reasonable and desired bound? To study such a concerning, let us consider that $U_\tau : [0,\infty) \to \mathbb{R}^d$ is the piecewise linear interpolant of the values of the sequence $(U_\tau^n)_{n\in\mathbb{N}}$, on the partition $\mathcal{P}_\tau$, defined by the relation

$$U_\tau(t) := \frac{t - t_\tau^{n-1}}{\tau} U_\tau^{n-1} + \frac{t_\tau^n - t}{\tau} U_\tau^n, \quad \text{for } t \in [t_\tau^{n-1}, t_\tau^n]. \tag{1.27}$$

The following result provides the convergence of the sequence $(U_\tau^n(t))_{n\in\mathbb{N}}$ to the solution of u_t to the problem (1.15) as $\tau \downarrow 0$.

Proposition 1.3.1. *If $U_\tau^0 \longrightarrow u_0$ as $\tau \downarrow 0$ then the family of piecewise interpolants $(U_\tau)_{\tau>0}$ satisfies the Cauchy condition as $\tau \downarrow 0$ with respect to the topology of uniform convergence on each compact interval $[0,T]$, for $T > 0$. Moreover, its unique limit is the solution u_t to* (1.15)*, and for every $T > 0$ there exists a constant $C(\lambda, L)$ such that*

$$\sup_{t\in[0,T]} |u_t - U_\tau(t)| \le |u_0 - U_\tau^0| + C(\lambda, T)|\nabla\phi(u_0)|\tau. \tag{1.28}$$

Proof. Shall we present a skertch of proof, although we refer the interested reader to e.g. Rulla (1996) for relevant results.

Without loss of generality, we consider that $\lambda = 0$. Then we can apply the non-expansive property of the map J_τ defined in (1.25), that is

$$|J_\tau(x) - J_\tau(y) \le |x - y|, \quad \text{for every } x, y \in \mathbb{R}^d. \tag{1.29}$$

Based on this property, we can derive the following uniform bound

$$\tau^{-1}|U_\tau^n - U_\tau^{n-1}| = |\nabla\phi(U_\tau^n)| \le |\nabla\phi(U_\tau^{n-1})|, \quad \text{for } n \ge 1. \tag{1.30}$$

So, under this perspective, we can also derive that

$$|U_\tau'(t))| \le \sup_{n\in\mathbb{N}} \tau^{-1}|U_\tau^n - U_\tau^{n-1}| = \tau^{-1}|U_\tau^1 - U_\tau^0| \le |\nabla\phi(U_\tau^0)|, \quad \text{for every } t \in [0,\infty)\backslash\mathcal{P}_\tau. \tag{1.31}$$

In addition, since $|\nabla\phi(U_\tau^0)| \longrightarrow |\nabla\phi(u_0)|$ as $\tau \downarrow 0$, we can observe that the family of piecewise interpolants $(U_\tau)_{\tau>0}$ satisfies uniformly a Lipschitz condition. Thus, it admits a subsequences which is converging uniformly to a Lipschitz map u in each compact interval $[0,T]$, as a concequence of Arzelá-Ascoli theorem (we will discuss it in the next chapter in detail). Moreover, if we denote $\overline{U}_\tau(t)$ the piecewise constant interpolant, that is

$$\overline{U}_\tau(t) := U_\tau^n, \quad \text{if } t \in (t_\tau^{n-1}, t^n], \tag{1.32}$$

playing the same game as in (1.31), we can derive

$$\sup_{t\in(0,\infty)} |U_\tau(t) - \overline{U}_\tau(t)| \le \tau|\nabla\phi(U_\tau^0) \tag{1.33}$$

Thus, we conclude that $\overline{U}_\tau(t)$ converges to the same limit as U_τ. In addition, thanks to our approximation scheme defined in (1.24), we have that

$$U_\tau' = -\nabla\phi(\bar{U}_\tau(t)), \quad \text{in } [0,\infty)\backslash\mathcal{P}_\tau \tag{1.34}$$

and passing to the limit in an integrated form, we can see that this u solves (1.15), which makes our skertch of proof complete. □

So, we can construct an effective approximation procedure, based on the classical Euler scheme, in order to "reach" the discrete solutions of gradient flow equation, and then, we can pass to the limit to obtain our genuine solutions. However, another, yet interesting and equivalent, way to see implicit Euler scheme is standing under the light of the famous *minimizing movements,* which was introduced by E. De Giorgi and his coworkers in 70s (see De Giorgi (1993); L. Ambrosio (1995)). The following remark presents this scheme.

Remark 1.3.2 (Minimizing Movements scheme)**.** Based on the functional $\Phi : (0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ defined in (1.26), one can define an equivalent variational scheme in order to obtain solutions of gradient flow type by solving recursively minimization problems. Precisely, for a given initial datum $u_0 \in \mathbb{R}^d$ and a time step parameter $\tau > 0$, we are looking for a sequence $(U_\tau^n)_{n\in\mathbb{N}}$ such that $U_\tau^0 := u_0$ and

$$\Phi(\tau, U_\tau^n, U_\tau^{n+1}) = \min_{U\in\mathbb{R}^d} \Phi(\tau, U_\tau^n, U), \quad \text{i.e.} \quad U_\tau^{n+1} \in \operatorname{argmin} \Phi(\tau, U_\tau^n, \cdot). \tag{1.35}$$

Under this perspective, any sequence satisfying (1.35), generates a discrete solution $U_\tau : [0,\infty) \to \mathbb{R}^d$, which is obtained at time step τ by a piecewise constant interpolation of values of the sequence $(U_\tau^n)_{n\in\mathbb{N}}$, i.e.

$$U_\tau(0) := U_\tau^0 = u_0 \quad \text{and} \quad U_\tau(t) := U_\tau(t) := U_\tau^n, \text{ for } ((n-1)\tau, n\tau], \text{ for every } n \in \mathbb{N}.$$

Based on these considerations, a curve $u : [0,\infty] \to \mathbb{R}^d$ is called *Minimizing Movement* related to Φ with initial value u_0 (denoted by $MM(\Phi, u_0)$) if the variational scheme (1.35) has a discrete solution U_τ converging to u as $\tau \downarrow 0$, that is

$$U_\tau(t) \to u(t), \qquad \text{as } \tau \downarrow 0, \text{ for } t \in (0,\infty).$$

In the same fashion, we can define the *Generalized Minimizing Movement* related to Φ with initial value u_0 (denoted by $GMM(\Phi, u_0)$) if there exists a suitable vanishing subsequence of time steps $(\tau_k)_{k\in\mathbb{N}}$ and corresponding discrete solutions U_{τ_k} at time step τ_k such that

$$U_{\tau_k}(t) \to u(t), \qquad \text{as } k \to \infty, \text{ for } t \in (0,\infty).$$

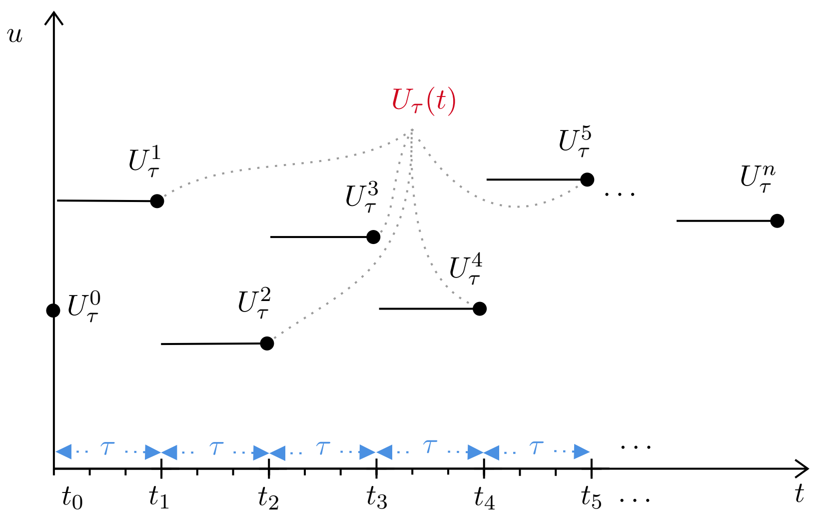

Figure 1.1: Minimizing Movements scheme.

In case of gradient flow type equations, a reasonable and natural choice for Φ, is described in (1.26). Using this fact, we can derive a variational formulation of implicit Euler scheme. Although, Minimizing movements is a generic concept, which can applied in a plethora of contexts, with different analytic, geometrical and topological flavor. Moreover, they will play a vital role to the construction of gradient flows in metric space setting.

1.4 Equivalent metric characterizations of gradient flows

At this point, we have already explored the notion of gradient flow in a Euclidean setting, and we discuss some fundamental estimates, which was presented in Proposition 1.2.4, under suitable assumptions of smoothness and convexity, which made us able to prove the well-posedness of gradient flows. Moreover, we have investigated several facts related with their numerical approximation, in the light of the implicit Euler scheme, and its variational twin, the Minimizing Movement scheme.

Having now, the fundamental estimates of Proposition 1.2.4 as stepping stone, we shall go a step further, by taking a closer look on them. The first interesting fact related with these estimates, is that there are not only estimates. Accurately, through them, we can obtain important and useful characterizations of gradient flows. Let us start the investigation of these characterizations of gradient flows by relaxing the Energy identity in (1.18). Doing this, we can arrive to an important characterization, which is described in the following proposition.

Proposition 1.4.1 (Energy Dissipation Inequality characterization)**.** *Consider that* $u : [0, \infty) \to \mathbb{R}^d$ *is* C^1 *map. Then* u *is a solution to* (1.15) *if and only if satisfies the Energy Dissipation Inequality (EDI):*

$$\frac{d}{dt}\phi(u_t) \leq -\frac{1}{2}|\frac{d}{dt}u_t|^2 - \frac{1}{2}|\nabla\phi(u_t)|^2, \qquad in\ (0, \infty). \tag{1.36}$$

Proof. Let $u \in C^1(\mathbb{R}^d)$. Employing the chain rule we get

$$\phi(u_t) = \phi(u_0) + \int_0^t \Big\langle \nabla\phi(u_r), \frac{d}{dr}u_r \Big\rangle dr, \tag{1.37}$$

so applying the energy dissipation identity given in (1.18), we get

$$\frac{1}{2}\int_0^t |u_t' + \nabla\phi(u_r)|^2 dr = \frac{1}{2}\int_0^t \Big(\Big|\frac{d}{dr}u_r\Big|^2 + |\nabla\phi(u_r)|^2\Big) dr + \int_0^t \Big\langle \nabla\phi(u_r), \frac{d}{dr}u_r \Big\rangle dr \leq 0. \tag{1.38}$$

From this, we can infer that

$$\frac{d}{dr}u_r = -\nabla\phi(u_r) \qquad \text{a.e. for } r \in (0, 1).$$

Moreover, by arbitrariness of t and since and $u \in C^1(\mathbb{R}^d)$, we conclude to the fact u solves (1.15) and this makes our proof. □

Very often, maps satisfying (1.36) are known as *curves of maximal slope.* We refer to L. Ambrosio et al. (2008) for a detailed account. However, as one can observe the characterization of above result *did not* make any use of the λ-convexity assumption of ϕ, which was imposed in previous sections. The whole argument was relying on just a application of the chain rule, which was possible due to our smoothness assumption.

Moving beyond the characterization of Energy Dissipation Inequality, we can obtain a useful result which characterizes gradient flows based on the Evolution Variational Inequality. This fact is presented in the following proposition.

Proposition 1.4.2 (Evolution Variational Inequality characterization)**.** *Consider that $u : [0,\infty) \to \mathbb{R}^d$ is a C^1 map satisfying the Evolution Variational Inequality (EVI) defined by* (1.17)*. Then u is a solution to* (1.15)*.*

Proof. Consider the function defined as

$$z \mapsto \mathsf{d}(z) = \frac{1}{2}|z-v|^2 \quad \text{for any fixed } v \in \mathbb{R}^d.$$

By a standard application of the chain rule on d, we infer that

$$\Big\langle \frac{d}{dt}u_t, u_t - v\Big\rangle \leq \phi(v) - \phi(u_t) - \frac{\lambda}{2}|u_t - v|^2, \quad \text{for every } v \in \mathbb{R}^d \text{ and } t > 0. \tag{1.39}$$

Now, for $\epsilon > 0$ and $\eta \in \mathbb{R}^d$, we choose $v := u_t + \epsilon\eta$ and dividing by ϵ we obtain

$$-\Big\langle \frac{d}{dt}u_t, \eta\Big\rangle \leq \frac{1}{\epsilon}\Big(\phi(u_t + \epsilon\eta) - \phi(u_t)\Big) - \frac{\lambda\epsilon}{2}|\eta|^2, \quad \text{for every } \eta \in \mathbb{R}^d. \tag{1.40}$$

Letting $\epsilon \downarrow 0$, we obtain

$$-\Big\langle \frac{d}{dt}u_t, \eta\Big\rangle \leq \langle \nabla\phi(u_t), \eta\rangle, \quad \text{for every } \eta \in \mathbb{R}^d. \tag{1.41}$$

Thus, we conclude that $\frac{d}{dt}u_t = -\nabla\phi(u_t)$, which completes our proof. □

Note again that the whole proof of the characterization through the Evolution Variational Inequality was based on the chain rule and a clever "radial" trick. The λ-convexity of ϕ does not seem to play any key role in this situation too. Although, based on the Evolution Variational Inequality, we can achieve a nice implication related with λ-convexity of ϕ, as the following result presents.

Proposition 1.4.3. *Suppose that there exists a C^1-semigroup $(\tilde{\mathcal{S}}_t)_{t>0}$ of smooth maps $\tilde{\mathcal{S}}_t : \mathbb{R}^d \to \mathbb{R}^d$ such that the curve $u_t := \tilde{\mathcal{S}}_t(u_0)$ satisfies the Evolution Variational Inequality (EVI), defined in* (1.17)*. The functional $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is λ-convex.*

Proof. With no loss of generality, let us assume that $\lambda = 0$. We pick $u^a, u^b \in \mathbb{R}^d$ and we define $u^s := (1-s)u^a + su^b$ and $u_t^s := \tilde{\mathcal{S}}_t(u^s)$. Trivially, the λ-convexity of ϕ is equivalent to

$$\frac{d}{ds}\phi(u^s)\Big|_{s=0} \leq \frac{d}{ds}\phi(u^s)\Big|_{s=1}. \tag{1.42}$$

So, we only need to prove that (1.42) holds. To prove this, we will employ the definition of gradient flow and the Evolution Variational Inequality. We can obtain

$$\begin{aligned}
\frac{d}{ds}\phi(u^s)\Big|_{s=0} &= \langle \nabla\phi(u^a, u^b - u^a\rangle \overset{(1.15)}{=} -\Big\langle \frac{d}{dt}u_t^a\Big|_{t=0}, u^b - u^a\Big\rangle \\
&= \frac{d}{dt}\Big(\frac{1}{2}|u_t^a - u^b|^2\Big)\Big|_{t=0} \overset{(1.17)}{\leq} \phi(u^b) - \phi(u^a) \\
&\overset{(1.17)}{\leq} -\frac{d}{dt}\Big(\frac{1}{2}|ua - u_t^b|^2\Big)\Big|_{t=0} \overset{(1.15)}{\leq} \langle \nabla\phi(u_b, u^b - u^a\rangle \\
&= \frac{d}{ds}\phi(u^s)\Big|_{s=1},
\end{aligned}$$

which makes our proof complete. □

1.5 Extensions and the scope of thesis

Through out this introductory chapter, in our discussion of the *gradient flows* we have restricted ourselves to the simple case of the Euclidean setting, considering smooth functionals and employing the suitable properties of the topological structure of this setting. In an intuitive way, this was extremely useful, in order to understand the nature of gradient flows, which started by variational characterizations, like the *evolution variational inequality,* the *energy identity* and the *slope identity* as consequences of a relaxation of convexity using a technical trick and concluded by a suitable approximation using *implicit Euler scheme* based on semigroup theory in PDEs. We have to mention that many results, which are described in this chapter, based on the the fact that the setting under consideration was the complete, compact, finite dimensional Euclidean space setting. For that reason, this discussion on gradient flows *cannot* be translated, in general, to other more general settings. Nevertheless, many analogous theories have been developed during last four decades, in different and interesting directions. Nowadays, consists a very active research field being on interplay of Analysis, PDEs, Calculus of Variations, Optimal Transportation and Probability. In this section will describe some directions to extend this theory which is also a part of the scope of this thesis.

Another direction and possibly the first one that one can imagine as a generalization of the Euclidean setting, is to consider a gradient flow generated by a proper lower semicontinuous $\lambda-$convex functional $\phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, where $\mathcal{H}$ denotes a separable Hilbert space (possibly infinite dimensional). All of the estimates and results discussed in this chapter can be recasted replacing the gradient of ϕ by its subgradient defined by the multivalued operator

$$\partial\phi(x) := \{p \in \mathcal{H} : \phi(y) \geq \phi(x) + p(y - x), \text{ for all } x, y \in \mathcal{H}\} \tag{1.43}$$

using convex analysis techniques, which can be found in Hiriart-Urruty & Lemaréchal (2012); Rockafellar (2015), avoiding any strong compactness assumption. In this framework, the resolvent operator $J_\tau := (\mathsf{Id} + \tau\partial\phi)^{-1}$, as a refinement of the operator defined in (1.25), is single-valued and non expansive, i.e. satisfying

$$d(J_\tau(v), J_\tau(u)) \leq d(u, v), \quad \text{for } u, v \in \mathcal{H} \quad \text{and } \tau > 0 \tag{1.44}$$

where d stands for the metric induced by the inner product of $\mathcal{H}$. Based on these properties we can prove the convergence of the exponential formula

$$(J_{t/n})^n(u_0) \overset{n\to\infty}{\longrightarrow} u_t \quad \text{and} \quad d(u_{t,t/n})^n(u_0)) \leq \frac{2|\partial\phi(u_0)|t}{\sqrt{n}}, \tag{1.45}$$

and thus define a contraction semigroup of the proper domain of ϕ, i.e. $\overline{D}(\phi) = \{x \in \mathcal{H} : \phi(x) < \infty\}$, based on Crandall-Ligett generation argument. Under these considerations, the unique solution u_t can be characterized, as in Euclidean setting, by a *evolution variational inequality.* This formulation goes back to J.L.Lions and G. Stampacchia Lions & Stampacchia (1967). Moreover, an analoguous approximation technique, that has been done in Section 1.3, can be obtained (see e.g. Rulla (1996)). All of these tools and techniques can be placed within the name of *differential inclusions and maximal monotone operators in Hilbert spaces* which has been developed in 70s by H. Brezis, J.L.Lions, M.G. Crandall, T.M. Liggett and many others (see e.g. Lions (1969); Brézis (1971); Crandall & Liggett (1971)).

Another, and quite more general, direction to extend the theory of gradient flows, is to forget everything about the vector structure of ambient space and consider a lower semicontinous functional $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$, where $\mathcal{X}$ denotes a complete and separable metric space. The starting point of studying gradient flows in this setting is the *equivament metric characterizations,* which were mentioned in Section 1.4. In this spirit, combining techniques from Metric Geometry and Analysis in Metric spaces, one have to look for absolutely continuous curves satisfying the equivalent metric characterizations. Since in metric setting, any classical notions of differentiability collapsed, under the light of absolutely continuity, we are able to replace it with the metric analogies of it, like *metric derivative* and *metric slope.* This theory is based on local compactness assumptions on ϕ and a kind of other assumptions on its (metric) slope which lead us to achieve lower semicontinuity, providing us a enormous flexibility but also a great disadvantange in terms of general metrics spaces many uniqueness and stability remains still open open problems. This direction can be placed under the roof of *curves of maximal slope in metric spaces,* which developed in 80s by E. De Giorgi, De Giovanni, Marino and Tosques in series of works (see for e.g. De Giorgi E. & Tosques (1980); De Giorgi et al. (1983); L. Ambrosio (1995); Cardinali et al. (1997)), or more recently, in work of L. Ambrosio, N. Gigli and G. Savare (see L. Ambrosio et al. (2008)).

A third direction could be through the so-called *generalized minimizing movements,* introduced by E. De Giorgi (see De Giorgi (1993)). This direction provides a general approximation scheme which, based on compactness arguments, is constructing limit curves. It is appearing in a plethora of cases and provides an important toolbox to treat different situations (see e.g. Gianazza & Savaré (1996)). Perhaps, one of the most important result in this perspective is the seminal paper or R. Jordan, D. Kinderlehrer and F. Otto (see Jordan et al. (1998)), where has been introduced the currently well-known JKO schemes.

Having all of these directions in our mind, let us specify the scope of this thesis. Our aim is to present the metric theory of gradient flows almost from scratch, having as starting point the equivalent metric characterizations of gradient flows, and moreover, to present some important applications of them. To do this, we start this investigation by considering the notion of curves in metric spaces, starting from the classical analysis of Lipschitz maps between metric spaces. Moreover, we will proceed to lengths of curves and length spaces, and furthermore to geodesics, geodesic spaces and geodesic convexity. Additionally, we will present a kind of crash course on Optimal Transport, which can provide us with a robust way to metrize the space of probability measures and understand many qualitative properties of a plethora of objects of our interest. Equally, based on optimal transport problem, we will study the metric, topological and geometrical properties of the so-called Wasserstein spaces, which will provide us a powerful machinery to study and understand PDEs in metric setting, as well as the understanding the nature of optimal transport problem itself. The last scene of this investigation will be the exploration of gradient flows in metric setting, where we will discuss the relations of equivalent metric characterizations, the minimal assumptions of existence of solutions. Under this perspective, we will study the gradient flows for geodesically convex functional, since they allow us to present several chic results. The equivalent metric characterizations coupled with the metric nature of Wasserstein spaces will led us to the so-called *Wasserstein gradient flows.*

CHAPTER 2

Curves on metric spaces

"I would like to make a confession which may seem
immoral: i do not believe in Hilbert space anymore."
— *John von Neumann*,
Hungarian mathematician (1903-1957)

Trying to extent the classical theory of gradient flows in vector space setting to a pure metric setting, one will encounter several difficulties, since the key notions which was applied in vectorial setting might be meaningless.

In this chapter, we present the key ingredients which will make us able to study gradient flows in pure metric setting. We will start by warming up ourselves about maps between metric spaces, we will study the notion of length of curves in metric spaces, ending up to so-called *length spaces.* Moreover, based on length spaces, we will explore geodesics and geodesic spaces, presenting many of their properties, arriving to *geodesic convexity.* Almost all of the results which will follow can be found to Burago et al. (2001); Papadopoulos (2005).

2.1 Warm up: some facts about maps in metric spaces

Before the investigation of curves in metric spaces, we shall recall some facts from Analysis of Lipschitz maps in metric spaces.

Let us consider that the pair $(\mathcal{X}, d)$ is a metric space equipped with the Hausdorff (metric) topology τ_d, that is the set $\mathcal{X}$ is equipped with a functional $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, which satisfies the properties of a metric, i.e it is non-negative, symmetric, indiscernible and subadditive and the open sets of τ_d are defined by the open balls $B_r(x)$ with respect to distance d, for some $x \in \mathcal{X}$ and some radius $r > 0$.

Having the notion of metric spaces in our mind, maybe the first reasonable question to ask is whether, distance is preserved between two metric spaces. Or, in other words, whether could we treat different objects, which belong in different metric spaces as the same, with respect to the notion of distance. This concern lead us to define the so-called *isometries* or *isometric embeddings.*

Definition 2.1.1 (Isometric embedding)**.** Consider that $(\mathcal{X}, d_{\mathcal{X}}), (\mathcal{Y}, d_{\mathcal{Y}})$ are two metric spaces. A map $f : \mathcal{X} \to \mathcal{Y}$ is called isometric embedding if it holds

$$d_{\mathcal{Y}}(f(x), f(y)) = d_{\mathcal{X}}(x, y), \quad \text{for every } x, y \in X.$$

Perhaps, in order to relax the restrictive distance preserving property of isometries, the first thought is to allow contractions and dilations in the distance preserving relation between two metric space. Such a thought, led us to the notion, Lipschitz maps.

Definition 2.1.2 (Lipschitz map)**.** Consider that $(\mathcal{X}, d_{\mathcal{X}}), (\mathcal{Y}, d_{\mathcal{Y}})$ are two metric spaces. A map $f : \mathcal{X} \to \mathcal{Y}$ is called Lipschitz if for some $L \geq 0$, it holds

$$d_{\mathcal{Y}}(f(x), f(y)) \leq L d_{\mathcal{X}}(x, y), \quad \text{for every } x, y \in \mathcal{X}. \tag{2.1}$$

The minimal L in the condition (2.1) is usually referred as the *Lipschitz constant* and denoted as $\mathrm{Lip}(f)$, that is

$$\mathrm{Lip}(f) := \inf\{L : f : \mathcal{X} \to \mathcal{Y} \text{ is Lipschitz map }\}.$$

Based on the definition of Lipschitz maps, we can define the bi-Lipschitz maps, which present several nice properties, as we will see in the following.

Definition 2.1.3 (Bi-Lipschitz map)**.** Consider that $(\mathcal{X}, d_{\mathcal{X}}), (\mathcal{Y}, d_{\mathcal{Y}})$ are two metric spaces. A map $f : \mathcal{X} \to \mathcal{Y}$ is called bi-Lipschitz if there exists $L \geq 1$ such that

$$\frac{1}{L} d_{\mathcal{X}}(x, y) \leq d_{\mathcal{Y}}(f(x), f(y)) \leq L d_{\mathcal{X}}(x, y), \quad \text{for every } x, y \in \mathcal{X}.$$

An interesting fact about bi-Lipschitz maps is that they are homeomorphic onto their images. From this observation, it follows that if $\mathcal{X}$ is a complete metric space, $\mathcal{Y}$ is a metric spaces and $f : \mathcal{X} \to \mathcal{Y}$ is a bi-Lipschitz map, then $\mathcal{Y}$ should be complete too. This statement of course does not hold for general homeomorphisms between metric spaces, since they cannot preserve properties like completeness.

It is widely known that there are many important theorems which are built around Lipchitz maps. Maybe the three most important are the *Banach's fixed point theorem,* which is involving a special case of Lipschitz maps where $L \leq 1$, the so-called contractions, the *Rademacher's theorem,* which describes the "boundaries" of differentiability of Lipschitz maps, and the *Baire's theorem,* which has deep and unexpected consequences in functional analysis. The latter two will play a crucial role in our forthcoming discussion.

Although, in many situations working with Lipschitz maps, it is desired to get involved with extensions of them, in order to obtain a suitable approximation. Studying whether two metric spaces has such a Lipschitz extension property consists a topic of quite current research (see Jensen (1993) or Garrido & Jaramillo (2008)). Perhaps one of most classical extension result consists the *McShane-Whitney extension,* which is presented in the following.

Proposition 2.1.4 (McShane-Whitney extension)**.** *Consider that $(\mathcal{X}, d)$ is a metric space, a nonempty set $A \subset \mathcal{X}$ nonempty and a Lipschitz map $f : A \to \mathbb{R}$. Then, there exists a function $\tilde{f} \in \mathrm{Lip}(\mathcal{X}, \mathbb{R})$, with the same Lipschitz constant as f, which is the extension of f in $\mathcal{X}$, i.e. $\tilde{f}|_A = f$.*

Proof. Let us define for every $a \in A$ the map $f_a : \mathcal{X} \to \mathbb{R}$ by

$$f_a(x) := f(a) + L d_{\mathcal{X}}(a, x), \quad \text{for every } x \in X.$$

Clearly, by its construction f_a is L-Lipschitz. Moreover, let us define the function $\tilde{f}$ as

$$\tilde{f}(x) := \inf\{f_a(x) : a \in A\}, \quad \text{for every } x \in \mathcal{X}.$$

Again, by construction, we can see that $\tilde{f}(x) < +\infty$ for any $x \in \mathcal{X}$. Now, fixing a $a_0 \in A$, we have

$$\begin{aligned} f(a) + L d_{\mathcal{X}}(a, x) &\geq f(a) + L d_{\mathcal{X}}(a, a_0) - L d_{\mathcal{X}}(a, x) \\ &\geq f(a_0) - L d_{\mathcal{X}}(a_0, x), \end{aligned}$$

and hence, we have that $\tilde{f}(x) > -\infty$ for any $x \in \mathcal{X}$. Thusly, since any f_a is L-Lipschitz, we have that $\tilde{f}$ is L-Lipschitz.

The only thing to prove is that $\tilde{f}$ extends f. To prove such a claim, we observe that for any $x \in A$ we have

$$\tilde{f} \leq f_x(x) = f(x) \leq f(y) + Ld_{\mathcal{X}}(x,y) = f_y(x), \qquad \text{for every} y \in A.$$

Therefore, $\tilde{f}|_A = f$ and our proof is complete. □

The above result admits interesting generalizations in more general settings (like Kirszbraun theorem on Hilbert spaces, see Schwartz et al. (1969)), but for the purposes of our discussion, such generalization are far from our goals. Based on McShane-Whitney extensions, one can provied a very useful approximation trick, which is presented in the following remark.

Remark 2.1.5 (Reverse extension trick)**.** The function $\tilde{f}$ in above proposition is the *smallest* possible extension. The largest one extension, following this spirit, for a $x \in \mathcal{X}$, is given by

$$\hat{f}(x) := \inf_{y \in A} \big\{ f(y) + \ell d_{\mathcal{X}}(x,y) \big\}.$$

This observation will be crucial in the following, making us able to prove several desired properties of certain functionals.

Keeping the discussion in the level of extensions, Lipschitz maps, we can have also a nice behaviour with respect to dense subspaces of their domain, as one might expect. A kind of such a behaviour is presented in the following result.

Proposition 2.1.6 (Dense embedding)**.** *Consider that $(\mathcal{X}, d_{\mathcal{X}})$ is a metric space and $\mathcal{X}' \subseteq \mathcal{X}$ is a dense subspace. Assume that $(\mathcal{Y}, d_{\mathcal{Y}})$ is a complete metric space and the Lipschitz map $f : \mathcal{X}' \to \mathcal{Y}$. Then there exists a unique Lipschitz map $\tilde{f} : \mathcal{X} \to \mathcal{Y}$ such that $\tilde{f}|_{\mathcal{X}'} = f$ and* $\mathrm{Lip}(f) = \mathrm{Lip}(\tilde{f})$.

Proof. For every $x \in \mathcal{X}$, let us pick up a sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_n \in \mathcal{X}'$ for any $n \in \mathbb{N}$ and $x_n \to x$. Then the sequence $\big(f(x_n)\big)_{n\in\mathbb{N}}$ is Cauchy, since $\mathcal{Y}$ is complete. This means that

$$d_{\mathcal{Y}}(f(x_n), f(x_m)) \leq Ld_{\mathcal{X}}(x_n, x_m) \to 0, \qquad \text{as } n, m \to \infty.$$

Thus, $L = \mathrm{Lip}(f)$. Moreover, again since $\mathcal{Y}$ is complete, we sure that there exists $y \in \mathcal{Y}$ such $f(x_n) \to y$. So, we define $\tilde{f}(x) = y$ and we can see that $\tilde{f}$ is well-defined, since it does not depend on the choice of the sequence $(x_n)_{n\in\mathbb{N}}$.

We claim that $\tilde{f}$ is L-Lipschitz. To see this, let $x, y \in \mathcal{X}$ and consider the sequences $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ such that $x_n \to x$ and $y_n \to y$. Then, we have

$$\begin{aligned} d_{\mathcal{Y}}(\tilde{f}(x), \tilde{f}(y)) &= \lim_{n\to\infty} d_{\mathcal{Y}}(f(x_n, f(y_n)) \\ &\leq L \lim_{n\to\infty} d_{\mathcal{X}}(x_n, y_n) \\ &= Ld_{\mathcal{X}}(x,y). \end{aligned}$$

The uniqueness, of such a function $\tilde{f}$ comes from the fact that if there were two continuous maps coinciding on a dense set, they have to coincide everywhere. This consideration makes our proof complete. □

Isometrical and dense embeddings can describe many important properties of metric spaces. One of the most important of them is that every metric space can be isometrically embedded to a complete metric space in a dense way. The procedure is the so called *metric completion.* In particular, we have the following result, which proof can be found in several Analysis books (see e.g. Bachman & Narici (2000)).

Proposition 2.1.7 (Completion of a metric space)**.** *Consider that $(\mathcal{X}, d)$ is a metric space. Then there exists a complete metric space $\tilde{\mathcal{X}}$ and an isometric embedding $f : \mathcal{X} \to \tilde{\mathcal{X}}$, such that the image $f(\mathcal{X})$ is dense in $\tilde{\mathcal{X}}$. The space $\tilde{\mathcal{X}}$ is unique up to any isometry and it is called completion of $\mathcal{X}$.*

Since now, we have carefully seen several important results related to the extensions, approximations and topological nature of Lipschitz maps. Through all of these results, several features of Lipschitz maps can be understood. Although, one may ask, what about the approximation of functions, using Lipschitz maps? Such a concern is very important problems in Calculus of Variations or Approximation theory. Perhaps, there is no unique answer, since it depends on the corresponding functional setting. Although, there is a very useful result, according to it any lower semicontinuous map can be viewed as a limit of an increasing sequence Lipschitz maps.

Proposition 2.1.8. *Consider that $(\mathcal{X}, d)$ is a metric space and $f : \mathcal{X} \to [c, \infty)$ is lower semicontinuous for some $c \in \mathbb{R}$. Then there exists a (increasing) sequence of Lipcschitz maps $(f_n)_{n\in\mathbb{N}}$ such that*

$$c \leq f_1 \leq \cdots \leq f_n(x) \leq f_{n+1}(x) \leq f(x) \quad \text{and} \quad f_n(x) \longrightarrow f(x), \quad \text{as } n \to \infty.$$

Proof. In case where $f(x) = +\infty$, we might choose $f_n(x) = n$ and the desired result will immediately follow.

In case where $f(x) < +\infty$, let us define for each $n \in \mathbb{N}$ the n-Lipschitz function f_n by

$$f_n(x) := \inf\{f(y) + nd_{\mathcal{X}}(x, y) : \quad y \in \mathcal{X}\}.$$

By construction, we can see that $c \leq f_1 \leq \cdots \leq f_n \leq f_{n+1} \leq f(x)$ for every $x \in \mathcal{X}$. Now, fixing $x \in \mathcal{X}$ and letting $M \in [c, f(x))$, we can choose a radius $r > 0$, such that $f(x) > M$ for every $x \in B_r(x)$. Then we have that $f_n(x) \geq \min\{M, c + nR\}$, so if we choose $n \in \mathbb{N}$ large enough, such that $c + nr > M$, we have $f_n(x) \geq M$. Thusly,

$$f_n(x) \longrightarrow f(x), \quad \text{as } n \to \infty,$$

as desired. So our proof is complete. □

We introduce now the notion of *uniform equicontinuity.*

Definition 2.1.9 (Uniform equicontinuity)**.** Consider that $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ are two metric spaces and $(f_n)_{n\in\mathbb{N}}$ is a sequence of maps $f_n : \mathcal{X} \to \mathcal{Y}$. We will say that $(f_n)_{n\in\mathbb{N}}$ is uniform equicontinuous if for any $\epsilon > 0$ there exists $\delta > 0$ such that for any $n \in \mathbb{N}$ and any $x, y \in \mathcal{X}$ we have

$$d_{\mathcal{X}}(x, y) < \delta \Longrightarrow d_{\mathcal{Y}}(f_n(x), f_n(y)) < \epsilon.$$

But, what about uniform equicontinuous maps between metric spaces? Why do we bother with such objects and which is the standard example of uniform equicontinuous maps? All of these sensible questions have specific answers, which will naturally appearing in the following. Let us only mention at this point that, thanks to uniform equicontinuity, is closely related with compactness-type arguments, which can be extremely useful in several situations. A standard example of uniformly equicontinuous sequence of maps, is a sequence of L-Lipschitz maps.

At this point, let us remind the notion of *diameter* in metric space. For a metric space $(\mathcal{X}, d)$, we define the its diameter as

$$\operatorname{diam}(\mathcal{X}) := \sup\{d(x, y) : x, y \in \mathcal{X}\}.$$

Based on diameter, we can define bounded subsets of $\mathcal{X}$ as the subsets with finite diameter with respect to the metric topology induced on these subsets. In order to continue this exploration, of the interplay of limits of sequence of maps between metric spaces and compactness-type arguments, we restrict ourselves to *proper metric spaces,* which are defined in the following.

Definition 2.1.10 (Proper metric space)**.** Consider that $(\mathcal{X}, d)$ is a metric space. We will say that $(\mathcal{X}, d)$ is proper if every bounded subset of $\mathcal{X}$ is compact.

The above definition reads to the fact that we can extract a convergent subsequence from any (infinite) bounded sequence. Under this prespective, proper metric spaces satisfies an analogue property like Heine-Borel property in any finite dimensional normed space. In addition, one can easily observe that a closed subset of a proper metric space, equipped with the induced metric topology is proper, and every proper metric space is complete. Moreover, a proper metric space is also separable space, i.e. it contains countable dense subset. All of these nice properties, turn proper metric space into a very nice and reasonable functional setting to work with, which can be extremely useful in many applications, in which we can characterize the limits of desired sequences of functionals describing a certain problem.

Perhaps, the most famous result of uniform equicontinuous maps in proper metric spaces is the *Arzelá-Ascoli theorem*, which characterize the boundedness of sequence of equicontinuous maps, and make us able to play the game of extraction of subsequences.

Theorem 2.1.11 (Arzelá-Ascoli)**.** *Consider that $(\mathcal{X}, d_{\mathcal{X}})$ is a separable metric space, $(\mathcal{Y}, d_{\mathcal{Y}})$ is proper metric space and $(f_n)_{n\in\mathbb{N}}$ is a sequence of uniformly equicontinuous maps $f_n : \mathcal{X} \to \mathcal{Y}$, which is bounded for every $x \in \mathcal{X}$. Then there exists a subsequence of $(f_n)_{n\in\mathbb{N}}$ which converges uniformly on every compact subset of $\mathcal{X}$ to a uniformly continuous map $f : \mathcal{X} \to \mathbb{R}$.*

2.2 Curves in length spaces

Having warmed ourselves up about some general facts of maps between metric spaces, we shall focus now on more specific questions related to such maps. All of such questions will lead us to several very interesting geometric feautures of maps between metric spaces. These features can be put under the name *Metric Geometry,* which consists nowadays a quite active research field and presents unexpectected applications to many disciplines. Most of the following results can be found in Burago et al. (2001) and Papadopoulos (2005). We will start this discussion with specific and concrete example.

Imagine that we are sitting in a summer night on roofgarden of Hotel Grande Bretagne in Athens, enjoying our rum fashion cocktail. If we ask ourselves what about the distance between our table and another one in the roofgarden, the answer can be easily and intuitively occur. More precesily, using a minimal ammount of abstraction and recalling the the Euclidean space setting, the distance can be immediately (and *locally*) measured and assigned with a specific number. Although, sitting in the roof and watching the view of Parthenon, such a concern could be realised with a more general perspective: *what about the distance between our table and the Parthenon?* Of course, in the same fashion as the distance between tables, such a distance can be measured as a *straight line* using a kind of optical device. Nevertheless, this strategy of measurement seems to be extremely useless, since this distance can going through only by birds[1]. So, given that we are human beings, the distance between our table can be viewed as *path or curve,* which we can going through by many ways, where its our is assigned with a specific length. Under this perspective, we would ideally want minimize the length of the path, in order to enjoy the view of Athens from the Parthenon, as fast as possible.

This heuristic (and maybe abuse) example contains a very clear mathematical precept. To be more accurate, in many cases we need to start with lengths of a class of paths, and then, based on the class of the paths, derive some kind of distance. Living locally in the Euclidean space setting, as we were in the distance between two tables, we could easily measure a distance with the

[1] Perhaps, not even by birds!

usual Euclidean metric. However, what about measuring distance as a shortest path between two points, as in the case of distance between our table and the Parthenon? Perhaps, the most natural thought, is to introduce a new distance which can measure the shortest path between these two points. Mathematically speaking, we could say that this distance stands for an intrinsic metric, where the distance between two points is constructed by the minimal connecting these points, or more precisely, the infimum of the lengths of the paths between them.

So, the starting point of such a discussion might be the notions of a *path or curve* in a metric space $(\mathcal{X}, d)$. Let us use by now the terminology *curve,* instead of path. Under this consideration, in what will follow, as curves we mean maps of intervals to metric spaces, or accurately, we mean maps $\gamma : I \to \mathcal{X}$ defined on an interval $I \subset \mathbb{R}$. By an interval we mean any connected subset of real line which might be open or closed, finite or infinite or even singleton. For the purpose of our discussion, we will assume that $I = [a, b]$, where $a, b \in \mathbb{R}$ with $a \leq b$. We will usually refer the points a, b as *endpoints.*

Considering the above notion of curve, our first concern related to its length, how this curve is changing, or more accurate, what about a notion of variation of that curve, or in terms of straight lines, how similar is it to a piecewise linear curve. All of these concerns are captured in the notion of the so-called *total variation* of that curve.

Definition 2.2.1 (Total variation of a curve)**.** Consider that $(\mathcal{X}, d)$ is a metric space and $\gamma : [a, b] \to \mathcal{X}$. We define the total variation of γ in $[a, b]$ as the quantity

$$V_\gamma(a, b) := \sup\left\{ \sum_{i=1}^{n} |\gamma(t_i) - \gamma(t_{i-1})| : a \leq t_0 \leq \cdots \leq t_n \leq b \right\}.$$

Based on the definiton of total variation of a curve in some metric space, we can intuitevely define its length, as its maximal variation.

Definition 2.2.2 (Length of a curve)**.** Consider that $(\mathcal{X}, d)$ is a metric space and $\gamma : [a, b] \to \mathcal{X}$. We define the length of γ in $[a, b]$ as the quantity

$$\ell(\gamma) := \sup\left\{ V_\gamma(a, b) : \quad \text{for } [a, b] \subset I, \text{ where } I \text{ is an open interval } \right\}.$$

Moreover, we will say that the curve γ is *rectifiable* if $\ell(\gamma) < \infty$.

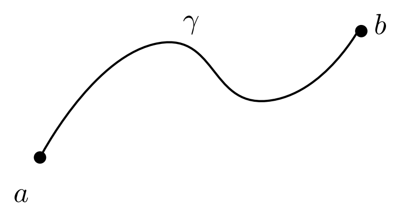

Figure 2.1: A curve γ connecting a with b

Having a well-defined notion of length for a curve defined in a metric space, we will focus on two classes of properties, which they enjoy. The first class of such properties are related to their parametrization. Reasonably, one may ask what about its behaviour along to changing of parameters. Before of proceeding to examine such a question, let us define clearly what we mean by *change of parameter* or *reparametrization.*

Definition 2.2.3 (Reparametrization)**.** Consider that $\gamma : [a, b] \to \mathcal{X}$ and $\gamma' : [c, d] \to \mathcal{X}$ are two curves in $\mathcal{X}$. We will say that γ' is obtained by γ through a reparemetrization if there exists a map $\psi : [c, d] \to [a, b]$ which is monotone, surjective and satisfies $\gamma' = \gamma \circ \psi$.

Note that, topologically speaking, we don't require that reparametrization map be homeomorphisms. Reparametrization of length of curves seems to be at this level very useful in many cases.

Maybe one of the most important properties of reparametrization is the *invariance of length.* Accurately, the length of a curve seems to be independent of the reparametrization, as it presented in the following result.

Proposition 2.2.4 (Length invariance under reparametrizations)**.** *Consider that $\gamma' : [c,d] \to \mathcal{X}$ be a curve obtained from a path $\gamma'[a,b] \to \mathcal{X}$ by reparametrization. Then the curves γ and γ' have the same length, that is $\ell(\gamma) = \ell(\gamma')$.*

Proof. We shall first prove the "greater or equal" direction. For that reason, let us consider that $\psi : [c,d] \to [a,b]$ is the change of parameter. We associate any partition $a \leq t_1 \leq \cdots \leq t_n \leq b$ of $[a,b]$, another partition $c \leq t'_1 \leq \cdots \leq t'_n \leq d$ of $[c,d]$, by choosing for each $i = 1, \cdots, n$, an arbitrary point in the set $\psi^{-1}(t_i)$ and re-ordering the points, if it is necessary. Then we have $V_{\gamma'}(c,d) \geq V_\gamma(a,b)$. Now, taking the supremum over all partitions of $[c,d]$, we obtain $\ell(\gamma') \geq V_\gamma(a,b)$. In the same fashion, taking the supremum over all partitions of $[a,b]$, we obtain

$$\ell(\gamma') \geq \ell(\gamma). \tag{2.2}$$

We shall prove now the "less or equal" direction. To do this, let us consider that P is a partition of $[c,d]$. Then, through ψ, we have that its image is $\psi(P) = P'$, where P' is a partition of $[a,b]$. Moreover, since ψ is monotone, we have that $V_{\gamma'}(c,d) = V_\gamma(a,b)$. Now, taking the supremum over all partitions of $[c,d]$ we obtain $V_{\gamma'}(c,d) \leq \ell(\gamma)$. In the same fashion, taking the supremum over all partitions of $[a,b]$, we obtain

$$\ell(\gamma') \leq \ell(\gamma). \tag{2.3}$$

Combining (2.2) and (2.3), we infer the invariance of length under parameter changing. □

Let us restrict ourselves to the case of rectifiable curves. In such a case, reparametrizations enjoy several thankful properties. We shall start the study of this case by recalling the following lemma. Its proof is ommitted, but in any case, we refer the interested reader to Burago et al. (2001).

Lemma 2.2.5. *Consider that $\gamma : [a,b] \to \mathcal{X}$ is a rectifiable curve. Then for every $u \in [0, \ell(\gamma)]$ there exists a unique $x \in \mathcal{X}$ and a $t \in [a,b]$ such that $x = \gamma(t)$ with $\ell(\gamma(t)) = u$.*

Based on above lemma, we could obtain several results related to analytical properties of reparametrizations. An example of such results is presented in the following proposition, which witness a Lipschitz property involved into reparametrization arguments.

Proposition 2.2.6. *Consider that $\gamma : [a,b] \to \mathcal{X}$ is a rectifiable curve. Let also the map $\zeta : [0, \ell(\gamma)] \to \mathcal{X}$ defined as $\zeta(u) = \gamma(t)$, where $\gamma(t)$ is the unique point according Lemma 2.2, satisfying $\ell(\gamma(t)) = u$. Then the map ζ is 1-Lipschitz.*

Proof. Let $u, u' \in [0, \ell(\gamma)]$ such that $u \leq u'$ and consider also $t, t' \in [a,b]$ such that $\ell(\gamma(t)) = u$ and $\ell(\gamma(t')) = u'$. Then, we have that $\zeta(u) = \gamma(t)$ and $\zeta(u') = \gamma(t')$. Considering the trivial partition $\{a, b\}$ of $[a,b]$, we observe that

$$d(u, u') \leq \ell(\gamma).$$

Using this fact, we have

$$\begin{aligned} d(\zeta(u), \zeta(u')) &\leq d(\gamma(t), \gamma(t')) \leq \ell(\gamma_{[t,t']} \\ &\leq \ell(\gamma(t') - \ell(\gamma(t)) \\ &= u' - u, \end{aligned}$$

whic shows that ζ is 1-Lipschitz and makes our proof complete. □

Having by now an clear view of reparametrization, we study an important notion of parametrizing curves, the so-called *arc length parametrization.* Intuitevely, this notion of parametrization is trying to imitate the case of "piecewise linearity" of a curve.

Definition 2.2.7 (Arclength parametrization)**.** Consider $\gamma : [a, b] \to \mathcal{X}$ is rectifiable curve. We will say that γ is parametrized by arclength if for every $a \leq u \leq v \leq b$, we have

$$\ell(\gamma|_{[u,v]}) = u - v.$$

Moreover, if $\gamma : [a, b] \to \mathcal{X}$ is parametrized by arclength it also holds that

$$\ell(\gamma) = b - a.$$

One could extend the notion of arclength parametrization in a homeomorphic way. Accurately, we have the following definition.

Definition 2.2.8 (Propotional to arclength parametrization)**.** Consider the curve $\gamma : [a, b] \to \mathcal{X}$. We will say that γ is parametrized propotionally to arclength if either γ is constant curve, or there exists a curve $\gamma' : [c, d] \to \mathcal{X}$ wich is parametrized by arclength and which satisfies $\gamma = \gamma' \circ \psi$, where $\psi : [a, b] \to [c, d]$ is the unique affine homeomorphism between these to intervals, that is the map defined as

$$\psi(x) := \frac{(d - c)x + (bc - ad)}{b - a}$$

Extending the notion of arclength parametrization of a curve through parametrization propotional to arclength one can again recast a Lipschitz property of that curve. In particular, as the following result witnesses, any propotional to arclength parametrization of a curve, uncovers to us a Lipschitz property with respect to the length of that curve.

Proposition 2.2.9. *Consider that $\gamma : [0, 1] \to \mathcal{X}$ is a curve parametrized propotionally to arc length. Then γ is $\ell(\gamma)$-Lipschitz map.*

Proof. Such a statement can be easily verified through the definition of parametrization propotional to arclength and the strategy of Proposition 2.2.6. For that reason, it is omitted. □

By this time, we have gently explored a class of important and significant reparametrization properties of length of curves. Now, we will focus on another class of crucial and desirable properties of them, that is *analytical properties,* which they are enjoying.

At the first level, it is no hard to see that, by construction, the length functional $\ell(\cdot)$, is *additive,* that is, for a path $\gamma : [a, b] \to \mathcal{X}$ and some $c \in [a, b]$ it holds that

$$\ell(\gamma) = \ell(\gamma|_{[a,c]}) + \ell(\gamma|_{[c,b]}).$$

Moreover, through close looking on the definition of the length functional, one might observe that for any $t \in [a, b]$, the map $t \mapsto \gamma(t)$ is *increasing and continuous.*

In order to work and explore properties of lengths of curves in a further analytical way, let us define the space of all curves in $\mathcal{X}$ with domain the interval $[a, b]$ as the set

$$\mathscr{C}([a, b], \mathcal{X}) := \{\gamma : [a, b] \to \mathcal{X} : \gamma(a) = x, \gamma(b) = y\},$$

equipped with the topology of uniform convergence, which is defined through the metric

$$d(\gamma_1, \gamma_2) := \sup_{t \in [a,b]} d(\gamma_1(t), \gamma_2(t)).$$

One may observe that the map $\ell : \mathscr{C}([a, b], \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ is not continuous in general. Although, it satisfies a lower semicontinuity property, which is presented in the following proposition.

Proposition 2.2.10 (Lower semicontinuity of length)**.** *Consider that* $(\mathcal{X}, d)$ *is a metric space. The length functional* $\ell : \mathscr{C}([a,b], \mathcal{X}) \to \mathbb{R} \cup \{+\infty\}$ *is lower semicontinuous.*

Proof. Let us fix a point $t \in [a,b]$. Then for any $\gamma_1, \gamma_2 \in \mathscr{C}([a,b], \mathcal{X})$ we have

$$d(\gamma_1(t), \gamma_2(t)) \leq d(\gamma_1, \gamma_2).$$

Therefore, the map $\gamma \mapsto \gamma(t)$ is continuous. Hence the map

$$\gamma \mapsto \sum_{i=1}^{n} |\gamma(t_i) - \gamma(t_{i+1})|$$

is continuous, as a sum of continuous maps. Therefore, the map

$$\ell(\gamma) = \sup\{V_\gamma(a,b) : \quad \text{for } [a,b] \subset I \text{ where } I \text{ is an open interval}\}$$

stands for a supremum of continuous functions, and hence it is lower semicontinuous. □

Another, and very useful, way to present the lower semicontinuity of length function $\ell(\cdot)$ is presented, in terms of sequences, in the following corollary. Its proof stands for a standard application of Proposition 2.2.10.

Corollary 2.2.11. *Consider that* $(\mathcal{X}, d)$ *is a metric space and* $(\gamma_n)_{n\in\mathbb{N}}$ *is a sequence of paths where* $\gamma_n : [a,b] \to \mathcal{X}$*, which converges to some path* $\gamma : [a,b] \to \mathcal{X}$*. Then we have*

$$\ell(\gamma) \leq \liminf_{n\to\infty} \ell(\gamma_n).$$

So by now, we have seen that lengths might not be continuous. Instead of continuity, we established a lower semicontinuity property. A reasonable concern to proceed our exploration on length of curves is related to sequences of length of curves and whether could we work with their subsequences, in order to describe desirable properties. Luckily, such questions, restricting ourselves to the case of proper metric spaces and thanks to the Arzelá-Ascoli theorem, can be easily treated and reduced to the discussion of the previous section. Accurately, we have the following result.

Proposition 2.2.12. *Consider that* $(\mathcal{X}, d)$ *is a proper metric space,* $L \geq 0$ *and for* $n \in \mathbb{N}$ *let* $\gamma : [0,1] \to \mathcal{X}$ *be a curve which is parametrized propotional to arclength such that* $\ell(\gamma) \leq L$*. Suppose also that the set* $\{\gamma_n(0) : n \in \mathbb{N}\}$ *is bounded. Then the sequence* $(\gamma_n)_{n\in\mathbb{N}}$ *has a subsequence which convergences to some curve* $\gamma : [a,b] \to \mathcal{X}$*, such that* $\ell(\gamma) \leq L$*.*

Proof. We have for any $n \in \mathbb{N}$ that γ_n is L-Lipschitz, so the sequence $(\gamma_n)_{n\in\mathbb{N}}$ is uniform equicontinuous. Moreover, since $\mathcal{X}$ is proper metric space and the sequence $(\gamma_n(0))_{n\in\mathbb{N}}$ is bounded, upon to a subsequence, we have $\gamma_n(0) \to \omega$, where $\omega \in \mathcal{X}$. Then, for every $n \in \mathbb{N}$ and for every $t \in [0,1]$ we have

$$\begin{aligned} d(\omega, \gamma_n(t)) &\leq \ell(\gamma_n) d(\gamma_n(a), \gamma_n(t)) \\ &\leq \ell(\gamma_n) d(t, 0) \\ &\leq L. \end{aligned}$$

Thus, for any $t \in [0,1]$ the sequence $(\gamma_n(t))_{n\in\mathbb{N}}$ is bounded. Thanks to the Theorem 2.1.11 (Arzelá-Ascoli), up to a subsequence, $\gamma_n \to \gamma$ uniformly. To conclude, thanks to Corollary 2.2.11, we have

$$\ell(\gamma) \leq \liminf_{n\to\infty} \ell(\gamma_n) \leq L,$$

which makes our proof complete. □

Having explored these few, but fundamental, properties of lengths of curves, we have developed a standard machinery to obtain the most important result of this section. Literally speaking, we are able now to prove that the infimum in the definition of length is finite, or in the other words, there exists a curve with minimal length.

Proposition 2.2.13 (Existence of curves with minimal length)**.** *Consider that $(\mathcal{X}, d)$ is a proper metric space and suppose that there exists a rectifiable curve $\gamma : [a, b] \to \mathcal{X}$, joining two points $x, y \in \mathcal{X}$. Then, the exists a curve whose length is equal to the infimum of lengths of all curves joining x and y.*

Proof. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence of paths joining x to y such that $\ell(\gamma_n) \to s$, where s is defined as

$$s = \inf\{\ell(\gamma) : \gamma : [a, b] \to \mathcal{X} \text{ such that } \gamma(a) = x, \gamma(b) = y\}.$$

Without loss of generality, we assume that for any $n \in \mathbb{N}$ is parametrized proportionally to arclength, and so, its domain is the interval $[0, 1]$. Thanks to Proposition 2.2.12, keeping the same notation up to a subsequence we have that $\gamma_n \to \gamma$. Taking the limits, we have clearly that the path γ joins x to y.

In addition, thanks to Corollary 2.2.11, we have

$$\ell(\gamma) \leq \liminf_{n\to\infty} \ell(\gamma_n) \leq s. \tag{2.4}$$

Moreover, by definition of s, we have that $\ell(\gamma) \geq s$. Combining this fact with (2.4), we obtain $\ell(\gamma) = s$, as it was desired. □

Having a well-defined notion of length in our toolbox, we can define a metric associated with that length notion, the so-called *intrinsic metrics.* Constructions like this, are motivate the definition of *length spaces.*

Definition 2.2.14 (Length space)**.** Consider that $(\mathcal{X}, d)$ is a metric space. Then for any two points $x, y \in \mathcal{X}$ we define the associated metric between them $d_\ell(x, y)$ as

$$d_\ell(x, y) := \inf\{\ell(\gamma) : \gamma : [a, b] \to \mathcal{X}, \quad \gamma(a) = x, \gamma(b) = y\}.$$

We will called d_ℓ intrinsic metric. Moreover, if $d = d_\ell$, the space $(\mathcal{X}, d_\ell)$ is called length space.

Proposition 2.2.15. *Consider that $(\mathcal{X}, d)$ is a length space, $x \in \mathcal{X}$ and $r > 0$. Then for every $y, z \in B_r(x)$, there exists a curve $\gamma : [a, b] \to \mathcal{X}$ with $\ell(\gamma) < 2r$ joining y and z. Moreover, the image of any such curve is contained to the $B_{2r}(x)$.*

Proof. By a standard application of the triangle inequality in $\mathbb{R}$ we have

$$|y - z| \leq |y - z| + |z - x| < 2r.$$

In addition, since $\mathcal{X}$ is a length space, there exist a path $\gamma : [a, b] \to \mathcal{X}$ such that $\ell(\gamma) < 2r$ joining y and z.

We will prove now that the image of such a curve is contained in B_{2r}. For the sake of contradiction, we suppose that there exists $t \in [a, b]$ such that $\gamma(t) \notin B_{2r}(x)$. Then we have

$$|y - \gamma(t)| \geq |x - \gamma(t)| - |x - y| > r \quad \text{ and } |z - \gamma(t)| \geq |x - \gamma(t)| - |x - z| > r.$$

Thus, we obtain

$$\ell(\gamma) = \ell(\gamma|_{[a,t]}) + \ell(\gamma|_{[t,b]}) \geq |y - \gamma(t)| + |z - \gamma(t)| > 2r,$$

which is a contradition, since $\ell(\gamma) < 2r$. This fact makes our proof complete. □

The following result, due to H.Hopf and W.Rinow, provides a nice characterization of compact subsets of completely locally compact length spaces, as an imitation of compact subsets of Euclidean space. We omit its proof, and we refer the interested reader to Gromov (2007).

Theorem 2.2.16 (Hopf-Rinow I). *Consider $(\mathcal{X}, d)$ that is complete and locally compact length space. Then $\mathcal{X}$ is proper metric space.*

Let us now make some technical comments on the definition of length space. One may note that dropping the assumption of proper metric space in above definition, the metric d_ℓ can take the value infinity. For example, if $\mathcal{X}$ is a disconnected union of two components, there is no continuous curve going from one component to other, and thus, the distance between two points in different components will be infinity. In addition, it might be no shortest curve between two points. For example, one can imagine the case of $\mathbb{R}^2$ where an open segment has been removed. In such a case, there is no shortest curve between the endpoints of removed segment. Although, the length can be still approximated, with a given precision, by other paths connecting endpoints.

However, under specific and reasonable assumptions, the structure of length space could a reasonable setting to work with. As one may expect, such structures stand on the intersection of Differential Geometry, Metric Geometry and Analysis in metric spaces, and nowadays, consist a pretty active research field (see e.g. the great monograph of M. Gromov (Gromov (2007)) or *the bible* of Metric Geometry of D. Burago et. al (Burago et al. (2001)) for an extended presentation) with many important applications.

For the purposes of this text, length spaces will be the basis in order to define and explore their children: *the geodesic spaces,* in the forthcoming discussion.

2.3 Absolutely continuous curves and differentiability

Having all of the notions of length of curves and their properties in our mind, let us make a "metric differential refuelling stop", in order to explore differentiability of curves in metric space setting, alongside the notions which were described in the previous section.

For the sake of doing this, a first observation can hint us that the study of such a concept in purely metric setting becomes meaningless. This happens because the classical notions of differentiability are strongly lying on the vectorial nature of the corresponding space, which in metric space has been disappeared

This fact motivates us, when one want to discuss mathematical phenomena which involving derivatives apart from vectorial setting, to find a suitable generalization of the "derivative" of a curve. The key idea to do this it is hidden, as usual, in the simplest setting which we can image, the Euclidean space setting. Thus, to recast such a notion of "derivative", we will focus to a specific question: *can we find an equivalent definition of derivative of smooth functions, using only the metric structure of the space?* Thankfully, the answer is positive, and stands in the core of notion of *absolutely continuity* of a curve.

We recall that the standard definition of absolute continuity of a curve $\gamma : [0, 1] \to \mathbb{R}$ suggests that for every $\epsilon > 0$, there exists $\delta = \delta(\epsilon)$ such that for every partition $\{0 = t_0 \leq t_1 < \cdots \leq t_n = 1\}$ of $[0, 1]$ it holds that

$$\text{if} \quad \sum_{i=1}^{n} |t_{i+1} - t_i| \leq \delta \quad \text{then} \quad \sum_{i=1}^{n} |\gamma(t_{i+1}) - \gamma(t_i)| \leq \epsilon.$$

Based on this classical definition we recall two important facts about absolutely continuous curves. The first one, as it is widely known, the absolutely continuity of a curve, thanks to a consequence of Rademacher Theorem. The second one stands in a very important and interesting observation, which is also essentially a corollary of its definition. To be more precise, as it can be proved, a curve is absolutely continuous *if and only if* there exist a function $g \in L^1([0,1])$, such that

$$|\gamma(t) - \gamma(s)| \leq \int_s^t g(r)dr, \quad \text{for every } s, t \in [0,1], \text{ with } s \leq t.$$

The main idea of extending the notion of absolute continuity of a function to a metric setting is hidden in the core of observation. This fact is witnessed by the following definition.

Definition 2.3.1 (Absolutely continuous curves)**.** Consider that $(\mathcal{X}, d)$ is a complete metric space and $\gamma : [a,b] \to \mathcal{X}$ is a curve. We will say that γ is absolutely continuous if there exists $g \in L^1([a,b])$ such that

$$d(\gamma(s), \gamma(t)) \leq \int_s^t g(r)dr \tag{2.5}$$

Moreover, we denote the space of all absolute continuous curves from $[a,b]$ to $\mathcal{X}$ by $AC([a,b]; \mathcal{X})$.

The above definition, can be naturally extended for functions $g \in L^p([a,b])$ and p-absolutely continuous curves for $p \in [1, +\infty)$,but such an extension is far from the purposes of this text.

Intuitively, one can observe that $AC([a,b]; X) \subset C([a,b]; X)$. The notion of absolute continuity is a *smoothness* property. In particular,i it s a weaker notion than continuity or uniform continuity, allowing us to obtain useful generalizations between the relation of integration and differentiation of curves in many situations. Moreover, all of Lipschitz maps are essentially absolute continuous.

An interesting consequence of above definitions is that for a given absolutely continuous curve, among all of the functions $g \in L^1([a,b])$ which satisfying (2.5), the minimal one is coincide with the modulus of the derivative of γ, which motivates the definition of *metric derivative.* Precisely, we have the following result.

Proposition 2.3.2 (Metric derivative)**.** *Consider that γ belongs to $AC((a,b); \mathcal{X})$. Then the limit*

$$|\gamma'|(t) := \lim_{s \to t} \frac{d(\gamma(s), \gamma(t))}{|s-t|}$$

exists for a.e. $t \in (a,b)$, and its called the metric derivative *of γ. Moreover, it is minimal in sense that*

$$|\gamma'|(t) \leq g(t), \quad \textit{for every } g \textit{ such as in } (2.5).$$

Proof. Consider a sequence $(x_n)_{n \in \mathbb{N}} \subset \mathcal{X}$, which is dense in the image of $[a,b]$ though γ. We define the function

$$d_n(t) = d(x_n, \gamma(t)), \text{ for every } t \in [a,b].$$

By construction, all of the terms $\{d_n : n \in \mathbb{N}\}$ are absolutely continuous in $[a,b]$, and therefore, the function

$$d(t) = \sup_{n \in \mathbb{N}} |d_n'|(t)$$

is well-defined a.e. in for a.e. $t \in (a,b)$. We choose a $t \in [a,b]$ where all of the terms d_n can be differentiated and we observe that

$$\liminf_{s \to t} \frac{d(\gamma(s), \gamma(t))}{|s-t|} \geq \sup_{n \in \mathbb{N}} \liminf_{s \to t} \frac{|d_n(s) - d_n(t)|}{s-t} = d(t). \tag{2.6}$$

But, since γ is absolutely continuous we have that (2.5) holds, and therefore, we obtain combining with (2.6), we infer that $d \geq g$, a.e. and hence $g \in L^1([a,b])$, which shows the minimality of $|\gamma'|(t)$.

Additionally, again thanks to construction and absolutely continuity of d, we have that

$$d(\gamma(s),\gamma(t)) = \sup_{n\in\mathbb{N}} |d_n(s) - d_n(t)| \leq \int_s^t d(r)dr$$

and thus

$$\limsup_{s\to t} \frac{d(\gamma(s),\gamma(t))}{|s-t|} \leq d(t). \tag{2.7}$$

Now, combining (2.6) and (2.7), the desired result follows and completes our proof. □

We shall now present two delicate examples, which can hint two distinct properties of metric derivative. Both of these properties witness its the power, which may be crucial in several situations

Example 2.3.3 (Coherency with classics)**.** The definition of metric derivative is coherent with the classical notion of derivative in vector space setting. To see this, let us assume for a moment that $\mathcal{X} = \mathbb{R}^d$. If a curve $\gamma : [a,b] \to \mathbb{R}^d$ is differentiable at some point t, then the metric derivative $|\gamma'|(t)$ is coincide with the norm of the classical derivative of γ at t.

Example 2.3.4 (Treating singluarities)**.** Another powerful property of metric derivative is that it can treat several singularities. We will see this by means of an example. Consider that $(V, \|\cdot\|)$ is a normed space, and d is the distance induced by its norm, i.e. $d(x,y) := \|x-y\|$ for $x, y \in V$. Then, for a given $v \in V$, we define the function

$$\gamma(t) := |t|v, \qquad \text{for } t \in [-1,1].$$

One can observe that the metric derivative of γ exists everywhere and it is coincide with its norm $\|\gamma\|$, but γ is not differentiable at $t = 0$, expect the trivial case where $v = 0$. Although, the metric derivative exists also at $t = 0$.

Taking into account the notion of absolute continuity of curves in metric spaces, as we saw, we passed through a suitable generalization of derivative, the metric derivative. So, having in our hands the stepping stone of differentiability, we can ask the question *could we define a "gradient-like" notion for functionals defined in pure metric setting too?* Hopefully, the answer is positive, and stands of fragments of the previous discussion about metric derivatives Thus, we have the following definition.

Definition 2.3.5 (Metric slope)**.** Consider that $(\mathcal{X}, d)$ is a metric space and $\phi : D(\phi) \subset \mathcal{X} \to \cup\{+\infty\}$ is a functional where $D(\phi) := \{x \in \mathcal{X} : \phi(x) < \infty\}$. Then for a given $x \in D(\phi)$, we define the metric slope of ϕ as

$$|\nabla\phi|(x) := \limsup_{y\to x} \frac{\big(\phi(x)-\phi(y)\big)^+}{d(x,y)} = \max\left\{\limsup_{y\to x} \frac{\big(\phi(x)-\phi(y)\big)^+}{d(x,y)}, 0\right\}.$$

2.4 Geodesics in metric spaces

Since now, we have described some basic facts about the lengths of curves in metric spaces. In addition we have already proceed to their notion of metric differentiability. As we discussed, we were able to prove that in a general metric setting, there exists a curve with minimal length joining two points of the corresponding space, that is the infimum of all curves joining two points is attained. This led us to the notion of length space. Moreover, living the length space space setting we saw that, thanks to Hopf-Rinow theorem, we can obtain a nice characterization of its compact sets.

Now, roughly speaking and based on this nice compactness-type characterization, as we will see in the following, we are able to prove that the above infimum luckily can be minimum. This procedure led us to the notion of curves with shortest length or *geodesics* in metric spaces. As it widely known, the notion of geodesics goes back to the era of F.W.Bessel and C.G Jacobi, and the science of geodesy, which is related to measurements of surface of the Earth and preexists of the modern mathematical treament of notion of length, which we presented in previous section. Although, one can treat geodesics as children of lengths of curves in a reasonable but also techical way.

To make the discussion more concrete, let us define the geodesics in metric setting.

Definition 2.4.1 (Geodesics and geodesic spaces)**.** Consider that $(\mathcal{X}, d)$ is a metric space and $\gamma : [a, b] \to \mathcal{X}$ is a curve. We will say that γ is a geodesic if

$$d(\gamma(x), \gamma(y)) = |x - y|, \qquad \text{for every } x, y \in [a, b]. \text{ with } x \neq y$$

Moreover $(\mathcal{X}, d)$ is called geodesic space if for every $x, y \in \mathcal{X}$ there exists a geodesic joining them.

As one can observe, geodesics are injective, and moreover, their restriction to a closed sub-interval of their domain remains geodesic. In addition, the above definitions reads as an isometric embeddings of the domain of the curve to its image. This definition is sightly different and more restrictive than the classical one in Riemannian geometry. Although, for the purposes of this text, seems to be pretty enough. As an alternative way to see geodesic space, we can say that $(\mathcal{X}, d)$ is a geodesic space if for any $x, y \in \mathcal{X}$ we have that

$$d(x, y) = \min\{\ell(\gamma) : \gamma : [a, b] \to \mathcal{X} \text{ joining } x \text{ and } y\ \}$$

Let us now a relevant notion of geodesics, that it the *constant speed geodesics*, which will be very important in the following.

Definition 2.4.2 (Constant speed geodesic)**.** Consider that $(\mathcal{X}, d)$ is a metric space. A curve $\gamma : [a, b] \to \mathcal{X}$ is called constant speed geodesic if there exists $K \geq 0$ such that

$$d(\gamma(x), \gamma(y)) = K|y - x|, \qquad \text{for every } x, y \in [a, b]. .$$

We will denote the metric space of all constant speed geodesics on $\mathcal{X}$ equipped with supremum norm, as $\mathscr{G}(\mathcal{X})$. An interesting fact is that the space $\mathscr{G}(\mathcal{X})$ inherits all of the topological properties of its base space $\mathcal{X}$ (see e.g. Carmo (1992)). Moreover, in case where $[a, b] = [0, 1]$, we recall the evaluation maps $e_t : \mathscr{G}(\mathcal{X}) \to \mathcal{X}$ defined for a geodesic $\gamma : [0, 1] \to \mathcal{X}$ and for every $t \in [0, 1]$ by

$$e_t(\gamma) := \gamma_t.$$

Having now the notion of geodesics on the table, maybe the first important fact about them is their parametrization. Thankfully, they can be parametrized by arclength, as the following result witness.

Proposition 2.4.3. *Consider that $(\mathcal{X}, d)$ is a metric space and $\gamma : [a, b] \to \mathcal{X}$ is a geodesic. Then γ is parametrized by arclength.*

Proof. Let $x, y \in \mathbb{R}$ such that $a \leq x \leq y \leq b$. Then, for any partition $a \leq t_1 \leq \cdots \leq t_n \leq b$ of $[x, y]$, we have

$$V_{\gamma|_{[x,y]}}(a, b) = \sum_{i=1}^{n} |\gamma(t_i) - \gamma(t_{i-1})| = \sum_{i=1}^{n} (t_i - t_{i-1}) = y - x.$$

Thus, we have

$$\ell(\gamma|_{[x,y]} = \sup\{V_{\gamma|_{[x,y]}} : \quad \text{for } [x, y] \subset I, \text{ where } I \text{ is an open interval}\} = y - x,$$

whichs shows that γ is parametrized by arclength and makes our proof complete. $\square$

Except their useful parametrization, geodesics are enjoying several other very interesting properties. The following proposition presents some of them related with its length and their behavior along the distance function. Such properties will be crucial in the following.

Proposition 2.4.4. *Consider that $(\mathcal{X}, d)$ is a metric space and γ is a curve parametrized by arclength. Then the following are equivalent:*

(i) *γ is geodesic*

(ii) *for any $x, y \in \mathbb{R}$ such that $a \leq x \leq y \leq b$, we have $d(\gamma(a), \gamma(y)) = d(\gamma(a), \gamma(x)) + d(\gamma(x), \gamma(y))$*

(iii) *$\ell(\gamma) = d(\gamma(a), \gamma(b))$*

Proof. $(i) \Rightarrow (ii)$: If γ is a geodesic, then, by its definition, there exist $x, y \in \mathbb{R}$ such that $a \leq x \leq y \leq b$ such that

$$d(\gamma(a), \gamma(y)) = y - a = y - x + x + a = d(\gamma(x), \gamma(y)) + d(\gamma(a), \gamma(x)),$$

which proves the desired conclusion.
$(ii) \Rightarrow (iii)$: Consider that $a \leq t_1 \leq \cdots \leq t_n \leq b$ is a partition of $[a, b]$. Applying n times the property (ii) we have

$$V_\gamma(a, b) = \sum_{i=1}^{n} |\gamma(t_i) - \gamma(t_{i-1})| = d(\gamma(a), \gamma(b)).$$

Moreover, taking the supremum of all partitions, we obtain

$$\ell(\gamma) = d(\gamma(a), \gamma(b)),$$

as it was desired.
$(iii) \Rightarrow (i)$: For any $x, y \in \mathbb{R}$ such that $a \leq x \leq y \leq b$, we have

$$\begin{aligned} \ell(\gamma) &= d(\gamma(a), \gamma(b)) \\ &\leq d(\gamma(a), \gamma(x)) + d(\gamma(x), \gamma(y)) + d(\gamma(y), \gamma(b)) \\ &\leq d(\gamma(a), \gamma(x)) + \ell(\gamma|_{[x,y]}) + d(\gamma(y), \gamma(b)) \\ &\leq \ell(\gamma|_{[a,x]}) + \ell(\gamma|_{[x,y]}) + \ell(\gamma|_{[y,b]}). \end{aligned}$$

Hence, all of above inequalities are equalities, and we have for any $x, y \in [a, b]$

$$d(\gamma(x), \gamma(y)) = \ell(\gamma|_{[x,y]})$$

Moreover, since γ is parametrized by arclength, we have $\ell(\gamma|_{[x,y]}) = |x - y|$, and thusly,

$$d(\gamma(x), \gamma(y)) = |x - y|,$$

which proves that γ is a geodesic, and makes our proof complete. □

We shall know go a step further, exploring some fragments of the topological properties of geodesics. In order to start such a procedure we present result, which is similar with Proposition 2.2.15, and witnesses the local behaviour of geodesics.

Proposition 2.4.5. *Consider that $(\mathcal{X}, d)$ is a length space, $x \in \mathcal{X}$ and $r > 0$. If $y, z \in B_r(x)$ (resp. $y, z \in \overline{B_r(x)}$) and if $\gamma : [a, b] \to \mathcal{X}$ is a geodesic joining y and z, then the image of γ is contained to $B_r(x)$ (resp. contained to $\overline{B_r(x)}$).*

Proof. The proof is essentially baseed in similar arguments as the proof of Proposition 2.2.15, and for that reason is ommitted. □

Another desirable property of geodesics, it that the (pointwise) limits of sequences of them remains geodesics. As it is presented in the following result, this fact stands as a standard consequence of continuity of distance function d on the corresponding metric space.

Proposition 2.4.6 (Limits of geodesics are geodesics)**.** *Consider that $(\mathcal{X}, d)$ is a metric space and a sequence of geodesics (resp. constant speed geodesics) $(\gamma_n)_{n\in\mathbb{N}}$ such that $\gamma_n : [a, b] \to \mathcal{X}$. If $(\gamma_n)_{n\in\mathbb{N}}$ converges (pointwise) to a map $\gamma : [a, b] \to \mathcal{X}$, then γ is geodesic (resp. constant speed geodesic).*

Proof. We shall prove this result in case where γ is geodesic, since if γ is constant speed geodesic the main arguments of the proof are essentially the same.

For any $x, y \in [a, b]$ and any $n \in \mathbb{N}$ we have

$$d(\gamma_n(x), \gamma_n(y)) = |y - x|.$$

By continuity of distance d, we immediately obtain

$$\lim_{n\to\infty} d(\gamma_n(x), \gamma_n(y)) = d(\gamma(x), \gamma(y)) = |y - x|,$$

and thus, γ is geodesic. □

The nice pointwise limit behaviour of geodesics (as well as constant speed geodesics) predispose us with nice hints about the topological properties of geodesics. The following result advise us that the game of extraction of subsequences can be in our favour in this case too. This fact is essentially a consequence of Arzelá-Ascoli theorem.

Proposition 2.4.7. *Consider that $(\mathcal{X}, d)$ is a compact metric space and $(\gamma_n)_{n\in\mathbb{N}}$ is a sequence of constant speed geodesics such that $\gamma_n : [a, b] \to \mathcal{X}$. Then the sequence $(\gamma_n)_{n\in\mathbb{N}}$ has a subsequence which converges uniformly to a constant speed geodesic $\gamma : [a, b] \to \mathcal{X}$.*

Proof. For any $n \in \mathbb{N}$ we have that $\ell(\gamma_n) = d(\gamma_n(a), \gamma_n(b))$. Since $\mathcal{X}$ is compact, we have that $\operatorname{diam}(\mathcal{X}) < \infty$, and hence, the sequence $\big(\ell(\gamma_n)\big)_{n\in\mathbb{N}}$ is bounded from above by a constant, which is independent of n. Thanks to Proposition 2.2.12, we have that $(\gamma_n)_{n\in\mathbb{N}}$ has a convergent subsequence, and keeping the same notion, we infer that $\gamma_n \to \gamma$ uniformly. Moreover, thanks to Proposition 2.4.6, we have that γ is a constant speed geodesic. □

By now, we have developed the standard machinery for a first understanding of geodesics. As we have already advertised, intuitively speaking, any geodesic space seems to be length space too. Now, based on our updated toolbox, let us rigorously proof such a statement.

Proposition 2.4.8. *Consider that $(\mathcal{X}, d)$ is a geodesic space. Then it is also length space.*

Proof. Let two points $x, y \in \mathcal{X}$ and a geodesic $\gamma : [a, b] \to \mathcal{X}$ joining x and y. We have that $|a - b| = |x - y|$, and moreover, since γ is geodesic, thanks to Proposition 2.4.3, is parametrized by arclength. So we have,

$$\ell(\gamma) = |a - b|.$$

Thusly, we obtain that $\ell(\gamma) = |x - y|$, which proves that $\mathcal{X}$ is a length space. □

Although, is the opposite statement true? Or in other words, does any length space be a geodesic space? Generally, the answer is negative. Although, restricting ourselves to the case where we can recast compactness-type arguments, we could say that the answer is positive. Let us start by presenting the following result, which stands for a local approach of existence of geodesics in length spaces.

Proposition 2.4.9 (Local existence). *Consider that $(\mathcal{X}, d)$ is locally compact length space. Then every $x \in \mathcal{X}$ has a neighborhood $\mathscr{U}$, such that for every $y, z \in \mathscr{U}$ there exists a geodesic in $\gamma : [a, b] \to \mathcal{X}$ joining y and z.*

Proof. We pick a point $x \in \mathcal{X}$. Since $\mathcal{X}$ is locally compact, there exists a radius $r > 0$ such that the ball $B_r(x)$ has a closure in $\mathcal{X}$. Let us define $\mathscr{U} := B_r(x)$ and pick up $y, z \in \mathscr{U}$.

Since $\mathcal{X}$ is length space, there exists a sequence of curves $(\gamma_n)_{n\in\mathbb{N}}$ such that $\gamma_n : [a, b] \to \mathcal{X}$ joining y and z with $\ell(\gamma_n) \to |y - z|$ as $n \to \infty$. Hence, if we choose large enough n, we could have

$$\ell(\gamma_n) < 2r.$$

Moreover, thanks to Proposition 2.4.5, the image of γ_n is contained in $\overline{B_{2r}(x)}$. Since this closure is compact, it is complete, and therefore, thanks to Theorem 2.1.11 (Arzelá-Ascoli), we infer that there exists a subsequence $(\gamma_{n_k})_{k\in\mathbb{N}}$ which converges to a curve $\gamma : [a, b] \to \mathcal{X}$ joining y and z with $\ell(\gamma) = |y - z|$. Without loss of generality, we assume that γ is parametrized by arclength. Then, thanks to Proposition 2.4.6, we infer that γ is a geodesic. Moreover, thanks to Proposition 2.2.15, we have tha the image of γ is contained in $B_{2r}(x)$ □

However, the above result can be much more global, using the assumption that metric space satisfies the property that every bounded is compact, and moreover, using standard properties of geodesics. The following result supports such a statement and stands as an alternative version of famous Hopf-Rinow theorem.

Theorem 2.4.10 (Hopf-Rinow II). *Consider that $(\mathcal{X}, d)$ is a proper length space. Then for every $x, y \in \mathcal{X}$, there exists a geodesic $\gamma : [a, b] \to \mathcal{X}$ joining x and y.*

Proof. Thanks to Proposition 2.2.13, we are ensured that there exists a curve $\gamma : [a, b] \to \mathcal{X}$ such that $\ell(\gamma) = |x - y|$. Without loss of generality, let us assume that γ is parametrized by arclength. Then, thanks to Proposition 2.4.4, γ is geodesics and our proof is completed. □

2.5 Geodesic convexity

Since now, we have already explored many desirable properties of curves in metric spaces, which led us to the notion of geodesic spaces. In this section, we will close up the discussion about curves in metric spaces, presenting one of the most important properties of geodesics spaces, that is *geodesic convexity.*

As it widely known, living in vectorial setting, convexity plays crucial role in certain and quite different situations, both abstract or not setting. It provides us several important results, as for example the famous *Mazur's Theorem* in topological vector spaces, or several important intuitive understanding, such as the interaction of a certain functional in vector spaces with the differential structure of the corresponding space.

One can ask, could we generalize the notion of convexity for functional on geodesic spaces? In the other words, can we define a notion of convexity in a pure metric setting which is compatible with some geometrical assumptions? The answer is positive and belongs in the heart of the notion of *geodesic convexity.* Let us define now the notion of geodesically convex subset, imitating the classical notion in vector space.

Definition 2.5.1 (Geodesically convex subset)**.** Consider that $(\mathcal{X}, d)$ is a proper geodesic space and $A \subset \mathcal{X}$. We will say that A is geodesically convex if for every $x, y \in A$, the geodesic joining x and y is contained in A.

Of course, by above definition of geodesically convex subset, one can easily observe that if $A \subset \mathcal{X}$ is geodesically convex subset and $h : \mathcal{X} \to \mathcal{X}$ is an isometry, then the image $h(A)$ is also geodesically convex subset. Let us mention at this point that we ask the geodesic space to be proper for a specific reason. If it is not proper, then as one can imagine there could be more than one notions of geodesic convexity. Accurately, we can ask for any pair of points in the subset, *at least one* or *any* geodesic joining them, to be contained in that subset, which will lead us to different situations. For proper geodesic spaces, these two notions are luckily coincide and makes the study of geodesic convexity much more simpler.

Let us now study some of topological properties of of geodesically convex subsets. Perhaps, the first one which we can imagine is related with the behaviour of set-theoretic operations of geodesically convex subsets. The following proposition presents a quite interesting behaviour of them with respect to set-theoretic operations.

Proposition 2.5.2 (Unions and intersections of geodesically convex sets)**.** *Consider that $(\mathcal{X}, d)$ is a proper geodesic space. Then the union of any increasing family of geodesically convex subsets is geodesically convex. Moreover, tha intersection of a family of geodesically convex subsets is geodesically convex.*

Proof. The proof stands for a standard application of definitions, and for that reason is omitted. □

Having the nice behaviour of intersections of geodesically convex subsets, we are able to define a geodesic analogue of the classical notion of convex hull in vector spaces, as it presents the following definition.

Definition 2.5.3 (Geodesic convex hull)**.** Consider that $(\mathcal{X}, d)$ is a proper geodesic metric space and $A \subset \mathcal{X}$. We define the geodesic convex hull of A as the intersection of all of the geodesically convex subsets of X that contained A.

Another interesting topological property of geodesically convex subsets, is that their closure is still a geodesically convex subset. This fact uses essentially the proper geodesic space setting and stands as a standard application of Arzelá-Ascoli Theorem, as it presents the following proposition.

Proposition 2.5.4 (Closure of geodesically convex subsets)**.** *Consider that $(\mathcal{X}, d)$ is proper geodesic space and $A \subset \mathcal{X}$ is a geodesically convex subset. Then its closure $\overline{A}$ is geodesically convex subset.*

Proof. We consider two points $x, y \in \overline{A}$ and we pick up two sequences $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}} \subset A$ such that $x_n \to x$ and $y_n \to y$, as $n \to \infty$. Let also a sequence $(\gamma_n)_{n\in\mathbb{N}}$ where $\gamma_n : [0,1] \to \mathcal{X}$ be a constant speed geodesic joining x_n and y_n. Thanks to Theorem 2.1.11 (Arzelá-Ascoli), the sequence $(\gamma_n)_{n\in\mathbb{N}}$ admits a subsequence $(\gamma_{n_k})_{k\in\mathbb{N}}$ which converges to a constant speed geodesic $\gamma : [0,1] \to \mathcal{X}$ joining x and y. Moreover, for any $t \in [0,1]$, since $\gamma(t)$ is the limit of $(\gamma_{n_k}(t))_{k\in\mathbb{N}}$, is contained in $\overline{A}$. Then, the image of γ is the unique geodesic in $\mathcal{X}$ joining x and y, which proves our thesis. □

Having understood some basic and fundamental properties of geodesically convex subsets, we are ready to define the analogue notion of convexity for functionals defined on geodesic spaces. This notion, will portray a very important role in the study of gradient flows in metric spaces, as we will see in the following.

Definition 2.5.5 (Geodesically convex functional)**.** Consider that $(\mathcal{X}, d)$ is a geodesic space. We will say that a functional $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is geodesically convex if for every $x_0, x_1 \in \mathcal{X}$ there exists a constant speed geodesic $\gamma : [0,1] \to \mathcal{X}$ connecting these two points such that

$$\phi\big(\gamma(t)\big) \leq (1-t)\phi\big(\gamma(0)\big) + t\phi\big(\gamma(1)\big), \qquad \text{for } t \in [0,1].$$

Similarly, we will say that $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is λ-geodesically convex if

$$\phi\big(\gamma(t)\big) \leq (1-t)\phi\big(\gamma(0)\big) + t\phi\big(\gamma(1)\big) - \frac{\lambda}{2} t(1-t) d^2(\gamma(0), \gamma(1)) \text{ for } t \in [0,1] \text{ and } \lambda \in \mathbb{R}.$$

Notice that, in general, this definition is not equivalent with usual convexity along the geodesic, since we only compare intermediate points t to 0 and 1, and not to other points. However, in the case of uniqueness of geodesics, geodesic convexity and convexity along geodesics coincides. Moreover, under the light of uniqueness of geodesics, we need it only for points which belong to $D(\phi)$, and not in the whole $\mathcal{X}$.

CHAPTER 3

A glance on Optimal Transport

> "There are no norms. All people are exceptions to a rule that doesn't exist. "
>
> — Fernando Pessoa,
>
> *Portuguese poet and philosopher (1888-1935)*

In this chapter, we recall the fundamental principles of *Optimal Transport.* This field has caught the attention of mathematical community, both of pure and applied mathematicians[1], since it is related with deep mathematical results which undercover fascinating mathematical phenomena.

3.1 Origin: Monge meets Kantorovich

The origin of Optimal Transport theory trace roots back to Gaspard Monge, a great enginerer and mathematician, and the era of French revolution. Monge, in his famous treatise (see Monge (1781)) formulated the first version of what we call now *optimal transportation problem.* Consider that we want to move a specific amount of earth from a given area, called *déblai,* to a given equal area, called *remblai,* with the least amount of effort, i.e. minimizing some transportation cost. As one can easily see, this formulation of Monge's problem is completely general and unclear, in terms of that its key ingredients, like cost or mass, are not fully mathematical specified. Although, it has been very famous for many decades (see e.g. Cayley (1883) and Vershik (2013)). Essentially, it was the inside the heart of a general class if many problems arising in differential geometry and analysis, standing as a bridge between abstract mathematical universe and physical world. For its solution, Academy of Paris was offering a prize (see Darboux (1885)). It was some years later, when P. Appell claimed its solution in his treatise (see Appell (1887)).

In a modern mathematical language, Monge's problem can be formulated as following. Consider, two complete and separable metric spaces $\mathcal{X}$ and $\mathcal{Y}$. Let also $\mu \in \mathscr{P}(\mathcal{X})$ and , $\nu \in \mathscr{P}(\mathcal{Y})$ two probability measures which represent the mass on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Since there is no free

[1]Maybe, this distinction is quite problematic. Instead of this one can imagine people working in mathematical universe and people working in physical world's problems with mathematical universe. Under this consideration, yet interesting, as it mentioned in Santambrogio (2015), we could say about people working *on* Optimal Transport and people who working *with* Optimal Transport. The purposes and the goals of each direction is useful, nevertheless extremely different, but this distinction maybe seems to be more concrete.

lunch, in order to move a point $x \in \mathcal{X}$ to a point $y \in \mathcal{Y}$, we have to pay a cost defined by a non-negative measurable function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. In this setting, we want to transport $\mathcal{X}$ to $\mathcal{Y}$, with a map $\boldsymbol{T} : \mathcal{X} \to \mathcal{Y}$, by investing the minimal effort possible. Of course, for our problem in becomes meaningful, if we want this map $\boldsymbol{T}$ be a measurable function and naturally has to preserving the masses, i.e. $\nu(B) = \mu(\boldsymbol{T}^{-1}(B))$ for any Borel measurable set $B \subset \mathcal{Y}$, or equivalently ν is the *push-forward measure* of μ by T which we will denote as $\boldsymbol{T}_{\#}\mu = \nu$. Under these considerations, *Monge's problem* is formulated as the following minimization problem:

$$\min_{T} \left\{ \int_{\mathcal{X}} c(x, \boldsymbol{T}(x)) d\mu(x) : \quad \boldsymbol{T}_{\#}\mu = \nu \right\}. \tag{3.1}$$

Nevertheless, a closer look to Monge's problem reveals many structural and technical pathologies, which drive to failure any possible general solution. In particular, its ill-posesness can be understood by means of the following reasons.

1. A transport map which solves Monge's problem may not exists or any transport map can solve Monge's problem. We will see this fact by means of two examples.

   Firstly, imagine the case where $\mu = \delta_x$ for some $x \in \mathcal{X}$ and $\nu \neq \delta_y$ for some $y \in \mathcal{Y}$. Then if $\boldsymbol{T} : \mathcal{X} \to \mathcal{Y}$, then $\nu(\boldsymbol{T}(x)) < 1 = \mu(\boldsymbol{T}^{-1}(\boldsymbol{T}(x))$ and one can see that there is no transport map! On the other hand, if one assume that $\mathcal{X} = [0, 1], \mathcal{Y} = [1, 2]$ and $c(x, y) = |x - y|$, using the linearity of integral in (3.1), we conclude that if $\boldsymbol{T}$ is a transport map then

   $$\int_{\mathcal{X}} c(x, \boldsymbol{T}(x) d\mu(x) \int_{\mathcal{X}} (\boldsymbol{T}(x) - x) dx = \int_{\mathcal{Y}} y dy - \int_{\mathcal{X}} x dx = 1 \tag{3.2}$$

   So, all transport maps gives the same value which is equal to 1.

2. The push-forward constraint $\boldsymbol{T}_{\#}\mu = \nu$ is highly nonlinear and **not** closed (and hence not compact) with respect any reasonable weak topology in the space of measures[2]. As a consequence, a limit of any minimizing sequence of transport maps $(\boldsymbol{T}_n)_{n \in \mathbb{N}}$ may fail to be a transport map. In other words, there is no analogue of Weierstrass theorem available in this setting, so the minimization becomes troublesome.

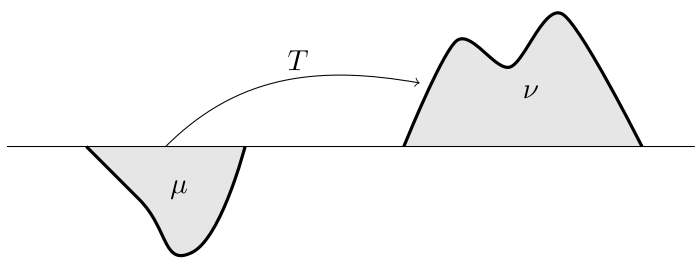

Figure 3.1: The transport of μ to ν through the map $\boldsymbol{T}$

In order to overcome these difficulties, 150 years later, Leonid Kantorovich proposed a nice and clever way to relax Monge's formulation (see Kantorovich (1942)). The main idea behind Kantorovich formulation is to focus on the fact that "graphs" of transport maps can be represented

[2] Of course, the space of measures can be equipped with the weak-* topology in L^∞ to obtain such a compactness argument. Although, this is not a reasonable choice of topology in the space of measures, since in this topology we cannot prove any lower semicontinuity property of the functional under minimization.

as probability measures on the product space. This is the key to a unlock limiting process, since the transportation problem can be reformulated in a space equipped with a reasonable weak topology of measures, and fill in enough compactness to construct such a minimizer. Technically speaking, in order to minimize over the set of all transport maps $\boldsymbol{T}$ satisfying the constraint $\boldsymbol{T}_{\#}\mu = \nu$, he proposed to solve the following minimization problem

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

where $\Pi(\mu,\nu)$ is the set of all admissible *transport plans,* i.e. probability measures $\pi \in \mathscr{P}(\mathcal{X}\times\mathcal{Y})$, having μ, ν as marginals, that is

$$\Pi(\mu,\nu) := \left\{ \pi \in \mathscr{P}(\mathcal{X}\times\mathcal{Y}) : \pi(A\times\mathcal{X}) = \mu(A), \quad \pi(B\times\mathcal{Y}) = \nu(A) \quad \text{for every } A \in \mathcal{B}(\mathcal{X}),\, B \in \mathcal{B}(\mathcal{Y}) \right\}$$

An equivalent formulation of the above set may be in terms of canonical projections $\mathsf{p}_{\mathcal{X}}, \mathsf{p}_{\mathcal{Y}}$ from the product space $\mathcal{X}\times\mathcal{Y}$ into $\mathcal{X}, \mathcal{Y}$ respectively, $(\mathsf{p}_{\mathcal{X}})_{\#}\pi = \mu$ and $(\mathsf{p}_{\mathcal{Y}})_{\#}\pi = \nu$.

Under this perspective, transport plans can be thought, in some sense, as "multivalued" transport maps since, for a given $x \in \mathcal{X}$, one can write $\pi = \int_{\mathcal{X}} \pi_x d\mu(x)$ with $\pi_x \in \mathscr{P}(\{x\}\times\mathcal{Y})$. Equivalently, one can say that for admissible transport plan π, the value of $\pi(A\times B)$ is the amount of mass which is sent from A to B. This formulation have several advantages, which makes it, more convenient to work with than Monge's. A few of them are presented in the following.

1. The set of all admissible transport plans $\Pi(\mu,\nu)$ is not empty (since trivially contains the tensor product $\mu\otimes\nu$) and enjoys some nice properties, such as convexity and compactness with respect to a suitable topology of convergence of measures as we will see in the following.

2. Under very minimal assumptions and using techniques of convex analysis, one can prove that mimimum always exists.

3. Transport plans include transport maps, since Monge's condition $\boldsymbol{T}_{\#}\mu = \nu$ implies that $\pi = (\mathsf{Id}\times\boldsymbol{T})_{\#}\mu$ lives in $\Pi(\mu,\nu)$.

We shall close this introductory section by presenting two very interesting and important examples of optimal transportation problems. The reason of doing this, is double-fold. In particular, on the one hand and as it will be enlightened in the following examples, optimal transportation problems can be interpreted in such a way, apprehending a useful and powerful machinery in abstract settings. On the other hand, thanks to their nature, they also can appropriate a very suitable framework to model unexpected applications, in many different disciplines e.g. statistics, economics or machine learning.

Example 3.1.1 (Continuous case: connections with Monge-Ampére equation)**.** Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and two probability measures $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ which are absolutely continuous with respect to d-dimensional Lebesgue measure $\mathscr{L}^d$. Thanks to Radon-Nikodym Theorem, we know that there exist two probability density function $f, g \in L^1(\mathbb{R}^d)$ such that $\mu = f(x)dx$ and $\nu = g(y)dy$. If $\boldsymbol{T} : \mathbb{R}^d \to \mathbb{R}$ is a smooth enough transport map, then the push-forward condition $\boldsymbol{T}_{\#}\mu = \nu$ can be re-written as a Jacobian equation.

More precisely, if $\phi : \mathbb{R}^d \to \mathbb{R}$ is any test function, then the push-foward condition yields to

$$\int_{\mathbb{R}^d} \phi\big(\boldsymbol{T}(x)\big) f(x)dx = \int_{\mathbb{R}^d} \phi(y)g(y)dy. \tag{3.4}$$

Assuming now that $\boldsymbol{T}$ is a diffeomorphism, if we set $y = \boldsymbol{T}(x)$ and using the change of variable formula, we can obtain that the right hand side of (3.4) is equal to

$$\int_{\mathbb{R}^d} \phi\big(\boldsymbol{T}(x)\big) g\big(\boldsymbol{T}(x)\big) \big| \det(\nabla \boldsymbol{T}(x)) \big| dx.$$

Thus, by arbitrariness of test function ϕ, we can get

$$\big| \det(\nabla \boldsymbol{T}(x)) \big| = \frac{f(x)}{g(\boldsymbol{T}(x))},$$

which can be recognized as *the Monge-Ampére equation.* On account of this observation, the solution of optimal transportation problem in terms of transport maps yields to the solution of Monge-Ampére equation. Under certain assumptions on cost function c, it is possible to recover many interesting bidirectional results, both in the theory of Monge-Ampére equation and the optimal transportation problem. A great treatise about this link and many of its further consequences can be found expository paper of De Philippis & Figalli (2014) and the recent monograph Figalli (2017).

Example 3.1.2 (Discrete case: optimal matching problem)**.** Now, let us restrict ourselves in the case where $\mathcal{X} = \{x_1, \cdots, x_n\}$, $\mathcal{Y} = \{y_1, \cdots, y_n\}$ are discrete spaces, $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a non-negative cost function and μ, ν are two empirical probability measures, giving uniformly to each point of each space the same probability, i.e.

$$\mu = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \quad \text{and } \nu = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}.$$

We can observe that any transport plan between μ and ν is represented by a $n \times n$ bistochastic matrix, $\pi = (\pi_{ij})_{i,j=1}^{n}$, where for every $j = 1, \cdots, n$ it holds that $\sum_i \pi_{ij} = 1$, and symmetrically, for every $i = 1, \cdots, n$ it holds that $\sum_j \pi_{ij} = 1$. In this case, Kantorovich problem is reformed to the following minimization

$$\inf \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_{ij} c(x_i, y_j) : \sum_{j=1}^{n} \pi_{ij} = 1 \text{ for every } i \text{ and } \sum_{i=1}^{n} \pi_{ij} = 1 \text{ for every } j \right\}, \tag{3.5}$$

which consists a infinite dimensional linear programming problem on the set of bistochastic matrices. Thanks to Choquet's theorem (see Phelps (2001) or Fabian et al. (2011)) and since the set of bistochastic matrices is convex, we understand that the problem (3.5) admits solutions which are on the extreme points of the set of bistochastic matrices. Although, thanks to Birkhoff's theorem (from which we know that any doubly stochastic matrix is a convex combination of permutation matrices, see L. Ambrosio et al. (2008)), these extreme points are a much simpler version of bistochastic matrices, that is, permutation matrices. Under these considerations, optimal transport plans in Kantorovich problems are exactly the same with solutions to Monge's problems, i.e.

$$\inf \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} c(x_1, y_{\sigma(i)}) : \sigma \text{ any permutation of } \{1, \cdots, n\} \right\}. \tag{3.6}$$

In many situations, problems like (3.6) are called *optimal matching problems.* On the other hand, this configuration of Kantorovich problem is important for two reasons. On the one hand, it paves the way to perform numerical implementations for computing the solution of an optimal transportation problem (see e.g. Peyré & Cuturi (2018) for a detailed description and many other perspectives), which may appeared in many several in economics, statistics or machine learning. On the other hand, it sheds the light to the reason of Monge's pathological formulation. To be more precise, in general, if two probability measures μ, ν are absolutely continuous with respect to Lebesgue measures, it may exists extreme points of $\Pi(\mu, \nu)$ which are not concentrated on any graph, so the call of Choquet's theorem is collapsed.

3.2 Existence of Optimal Transport plans

Let us focus now to the qualitative nature of optimal transport problem. Precisely, we will see that the minimizer of Kantorovich problem is really exists. In order to prove the existence of such a minimizer, we remind some and definitions propositions about probability measures in metric spaces. All of their proofs, instead of their mathematical beauty, are not the main purpose of this text and will be omitted. We refer the interested reader to Billingsley (2013) or Topsoe (2006) for a detailed description.

For what will follow, we consider that $(\mathcal{X}, d)$ is a complete and separable metric space, with Borel σ-algebra $\mathcal{B}(\mathcal{X})$, and with Hausdorff topology τ_d induced by its metric. We recall that the *support* of a propability measure $\pi \in \mathscr{P}(\mathcal{X})$, will be denoted as $\operatorname{supp}\mu \subset \mathcal{X}$, and defined as

$$\operatorname{supp}\mu := \overline{\{x \in \mathcal{X} : \mu(U) > 0 \text{ for each neighborhood } U \text{ of } x\}}. \tag{3.7}$$

We will say that a probability measure $\pi \in \mathscr{P}(\mathcal{X})$ is *tight,* if for any $\epsilon > 0$, there exists a compact set K_ϵ such that $\pi(K_\epsilon^c) < \epsilon$.

Since $\mathcal{X}$ is complete and separable metric space, we have that any $\pi \in \mathscr{P}(\mathcal{X})$ is regular, i.e. for a given set $A \in \mathcal{B}(\mathcal{X})$, it holds that

$$\mu(A) = \sup\{\mu(E) : E \subseteq A \text{ compact}\} = \inf\{\mu(\mathcal{O}) : \mathcal{O} \supseteq A \text{ open}\}.$$

In addition, again due to structure of $\mathcal{X}$, we have that any $\pi \in \mathscr{P}(\mathcal{X})$ is concentrated in a σ-compact set, i.e. there is a measurable set S, which can be written as the union of countably many compact sets, such that $\pi(S) = 1$ (see for details Billingsley (2013)). Another interesting and useful properties of probability measures on complete and separable metric spaces, is witnessed by famous Ulam's lemma, which is presented in the following proposition.

Proposition 3.2.1 (Ulam's lemma). *Consider that $(\mathcal{X}, d)$ is complete and separable metric space. Then any Borel probability measure $\pi \in \mathscr{P}(\mathcal{X})$ is tight.*

Now, turning our attention to convergence of probability measures, which will be crucial to study what will follows. For that reason, we recall the definition of *narrow convergence of measures.*

Definition 3.2.2 (Narrow convergence). Consider that $(\mathcal{X}, d)$ is a complete and separable metric space and $(\pi_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\mathcal{X})$ is a sequence of probability measures. We will say that $(\pi_n)_{n\in\mathbb{N}}$ converges to some $\pi \in \mathscr{P}(\mathcal{X})$ in narrow sense if

$$\int_{\mathcal{X}} f(x) d\pi_n(x) \longrightarrow \int_{\mathcal{X}} f(x) d\pi(x) \quad \text{ as } n \to \infty \tag{3.8}$$

for every function $f \in C_b(\mathcal{X})$, the space of continuous and bounded functionals defined on $\mathcal{X}$.

An interesting observation is that the space of probability measures $\mathscr{P}(\mathcal{X})$ can be identified with a convex subset of the unitary ball of the dual space of bounded and continuous functionals $\big(C_b\big)^*$. Indeed, by definition, narrow convergence is induced by the weak-* topology of $\big(C_b\big)^*$. Moreover, narrow convergence have several interesting and useful consequences. In the following proposition, we present some of them which will be in the background of many important results.

Proposition 3.2.3 (Portmanteau). *Consider that $(\mathcal{X}, d)$ is a complete and separable metric space and $(\pi_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\mathcal{X})$ is a sequence of probability measures. Then, the following are equivalent:*

1. $\int_{\mathcal{X}} f d\pi_n \to \int_{\mathcal{X}} f d\mu$, *for* $f \in C_b(\mathcal{X})$
2. $\limsup_n \pi_n(F) \leq \mu(F)$, *for every closed set* $F \subseteq \mathcal{X}$
3. $\liminf_n \pi_n(U) \geq \mu(U)$, *for every open set* $U \subset \mathcal{X}$.

Having the suitable notion of convergence of probability measures, we are able to recast several nice topological features, related to our problem. Perhaps, the most important of them is the equivalence of tightness and compactness. This fact is presented in the following theorem and it will play a fundamental role in what follows.

Theorem 3.2.4 (Prokhorov). *Consider that $\mathcal{X}$ is a complete and separable metric space. Then, if a collection of probability measures $K \subset \mathscr{P}(\mathcal{X})$ is tight, i.e. for any $\epsilon > 0$ there exists a compact set K_ϵ in $\mathcal{X}$ such that $\mu(\mathcal{X}\backslash K_\epsilon) \leq \epsilon$, for any $\mu \in K$, then K is relatively (sequentially) compact in $\mathscr{P}(\mathcal{X})$. Conversely, every relatively (sequentially) compact subset of $\mathscr{P}(\mathcal{X})$ is tight.*

As we have already mentioned, based on the compactness conclusion of Prokhorov's theorem, one can easily understand that provides a very useful tool to work, since it will allow us to play the game of extraction of subsequences to prove convergence with respect reasonable topologies. Now, since we recalled the necessary machinery, we can proof the first interesting property of the set of transport plants in the following proposition.

Proposition 3.2.5 (Transport plans are tight). *Consider two probability measures $\mu, \nu \in \mathscr{P}(\mathcal{X})$. Then the set of transport plans $\Pi(\mu, \nu)$ is tight.*

Proof. Let $\epsilon > 0$. By regularity of μ, ν, there exist compact sets $A \subset \mathcal{X}$ and $B \subset \mathcal{Y}$ such that $\mu(A) > 1 - \epsilon$ and $\nu(B^c) < \epsilon$. For some $\pi \in \Pi(\mu, \nu)$ we have

$$\begin{aligned} \pi(A \times B) &= \pi(A \times \mathcal{Y}) - \pi(A \times (B^c)) \\ &\geq \pi(A \times \mathcal{Y}) - \pi(\mathcal{X} \times (B^c)) \\ &= \mu(A) - \nu(B^c) \\ &> 1 - 2\epsilon \end{aligned}$$

Since $\pi \in \Pi(\mu, \nu)$ was arbitary and $A \times B$ is compact, we achieve tightness and our proof is completed.

□

Based on tightness of $\Pi(\mu, \nu)$, the existence of solution to Kantorovich problem comes from a standard consequence of lower-semicontinuity and compactness arguments.

Proposition 3.2.6 (Existence of minimizer for Kantorovich problem). *Consider that $\mathcal{X}, \mathcal{Y}$ are complete and separable metric spaces and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is lower-semicontinuous and bounded from below. Then, there exists a solution to the Kantorovich problem, as it stated in* (3.3).

Proof. Our strategy is to recall Weierstrass theorem for existences of minimizer. To do this, we will prove compactness of the set $\Pi(\mu, \nu)$ and the lower semicontinuity of the functional $\pi \mapsto \int_{\mathcal{X}\times\mathcal{Y}} c d\pi$.

We consider a minimizing sequence $(\pi_n)_{n\in\mathbb{N}} \subset \Pi(\mu, \nu)$. Since $\Pi(\mu, \nu) \subset \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ is tight, thanks to Proposition 3.2.5, due to Theorem 3.2.4 (Prokhorov's Theorem) is also relatively compact. Hence, there exists a subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ of $(\pi_n)_{n\in\mathbb{N}}$ such that $\pi_{n_k} \to \pi$ in narrow sense as $k \to \infty$.

We claim that $\pi \in \Pi(\mu, \nu)$. To see this, we take as test functions $\varphi(x, y) = \varphi_1(x)$, for $\varphi_1 \in C_b(\mathcal{X})$. Then, we note that

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi(x, y) d\pi_{n_k} \overset{k\to\infty}{\longrightarrow} \int_{\mathcal{X}\times\mathcal{Y}} \varphi(x, y) d\pi = \int_{\mathcal{X}} \varphi_1 d\mu,$$

while since $\pi_{n_k} \in \Pi(\mu, \nu)$ for every $k \in \mathbb{N}$, we have

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi(x, y) d\pi_{n_k} = \int_{\mathcal{X}} \varphi_1(x) d\mu.$$

Therefore, we have

$$\int_{\mathcal{X}} \varphi_1(x) d\mu = \int_{\mathcal{X}\times\mathcal{Y}} \varphi(x, y) d\pi(x, y),$$

which is equivalent to $(\mathsf{p}_{\mathcal{X}})_{\#}\pi = \mu$. In the same fashion, consider $\varphi(x, y) = \varphi_2(y)$, for $\varphi_2 \in C_b(\mathcal{Y})$. Similarly, we note that

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi(x, y) d\pi_{n_k} \overset{k\to\infty}{\longrightarrow} \int_{\mathcal{X}\times\mathcal{Y}} \varphi(x, y) d\pi = \int_{\mathcal{Y}} \varphi_2 d\nu,$$

while since $\pi_{n_k} \in \Pi(\mu, \nu)$ for every $k \in \mathbb{N}$, we have

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi(x, y) d\pi_{n_k} = \int_{\mathcal{Y}} \varphi_2(x) d\nu.$$

Thus. we have

$$\int_{Y} \varphi_2(y) d\nu = \int_{\mathcal{X}\times\mathcal{Y}} \varphi(x, y) d\pi(x, y),$$

thus we have $(\pi_{\mathcal{Y}})_{\#}\pi = \nu$.

Now, we wil prove the lower semicontinuity of functional $\pi \mapsto \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) d\pi(x, y)$. Since c is only lower semicontinuous, we need to employ an approximation argument. For that reason, we define for every $\ell \in \mathbb{N}$

$$c_\ell(x, y) = \inf_{(x', y') \in \mathcal{X}\times\mathcal{Y}} \Big\{ c(x', y') + \ell \mathsf{d}\big((x, y), (x', y')\big) \Big\},$$

where $\mathsf{d} : \big(\mathcal{X} \times \mathcal{Y}\big) \times \big(\mathcal{X} \times \mathcal{Y}\big) \to \mathbb{R}$ is a metric defined on $\mathcal{X} \times \mathcal{Y}$. We note that c_ℓ is ℓ-Lipchitz, while $c_\ell(x, y) \uparrow c(x, y)$, i.e.

$$c(x, y) = \sup_{\ell \in \mathbb{N}} c_\ell(x, y).$$

Thus, thanks to Monotone Convergence Theorem, for any $\pi \in \Pi(\mu, \nu)$, we have that

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x, y) d\pi = \sup_{\ell \in \mathbb{N}} \int_{\mathcal{X}\times\mathcal{Y}} c_\ell(x, y) d\pi.$$

We define the functional $\pi \mapsto \int_{\mathcal{X}\times\mathcal{Y}} c_\ell d\pi$, which is continuous with respect to topology of narrow convergence, thanks to continuity of c_ℓ. Hence we have that the functional $\pi \mapsto \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) d\pi$ is lower semicontinuous, since it can be written as supremum of a family of continuous functionals. Therefore, due to Weierstrass theorem, the problem (3.3) admits a solution, which makes our proof complete.

□

Let us comment some fact about the proof of above result. The latter approximation trick, undercover to us several properties of the cost functional c. In general, optimal transport problems are strongly related with the nature of their cost function. In the other words, we can prove several results for a specific class of cost functionals. However, the above approximation trick has an interesting extension for general cost functionals, which is presenting in the following result.

Proposition 3.2.7 (Lower semicontinuity of cost function)**.** *Consider that $\mathcal{X}, \mathcal{Y}$ are complete and separable metric spaces and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous cost function. Let also $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be and upper semicontinuous functional such that $c \geq h$. If $(\pi_n)_{n\in\mathbb{N}} \subset \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ is a sequence of probability measures converging to $\pi \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$, $h \in L^1(\pi_n)$, $h \in L^1(\pi)$ and*

$$\int_{\mathcal{X}\times\mathcal{Y}} h d\pi_n \to \int_{\mathcal{X}\times\mathcal{Y}} h d\pi, \qquad \text{as } n \to \infty,$$

then

$$\int_{\mathcal{X}\times\mathcal{Y}} c d\pi \leq \liminf_{n\to\infty} \int_{\mathcal{X}\times\mathcal{Y}} c d\pi_n.$$

Proof. We replace the cost function c by $c-h$ and we assume that $c \geq 0$, and still lower semicontinuous. Then, c can be written as a pointwise limit of a non decreasing family $(c_\ell)_{\ell\in\mathbb{N}}$ of continuous real-valued cost functions. Under this prospective, employing monotone convergence theorem, we have

$$\int_{\mathcal{X}\times\mathcal{Y}} c d\pi = \lim_{\ell\to\infty} \int_{\mathcal{X}\times\mathcal{Y}} c_\ell d\pi = \lim_{\ell\to\infty} \lim_{n\to\infty} \int_{\mathcal{X}\times\mathcal{Y}} c_\ell d\pi \leq \liminf_{n\to\infty} c d\pi_n,$$

which proves the desired result.

□

So, since now we are ensured that the minimizers of Kantorovich problem exists, by standard lower semicontinuity and compactness arguments. An interesting question arising after this fact is related to the structure of candidate minimizers and minimizers. In the other words, beyond existence, we are interesting about the structural properties of transport plans, as well as the optimal transport plan.

The first step to attack such concerns is to restrict ourselves to a simple (or sometimes not so simple) setting. Accurately, let us imagine the special case of Euclidean space setting with quadratic cost functional, that is $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and $c(x, y) = |x - y|^2$. Under this consideration, let $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ be two probability measures supported on finite sets. After a short calculation period, we could realize that a transport $\pi \in \Pi(\mu, \nu)$ is optimal if and only if it holds

$$\sum_{i=1}^{n} \frac{|x_i - y_i|^2}{2} \leq \sum_{1}^{n} \frac{|x_i - y_{\sigma(i)}|^2}{2}, \tag{3.9}$$

for any $n \in \mathbb{N}$, $(x_i, y_i) \in \operatorname{supp}(\pi)$ and σ a permutation of the set $\{1, \cdots, n\}$. Expanding the squares in (3.9), we obtain

$$\sum_{i=1}^{N} \langle x_i, y_i \rangle \leq \sum_{i=1}^{N} \langle x_i, y_{\sigma(i)} \rangle.$$

The relation described in (3.2) is nothing more but the definition of the notion of *cyclically monotonicity* which introduced in Convex Analysis by R.T Rockafellar in 70s (see e.g. Rockafellar (2015)). On the basis of this observation, we can generalize this notion to c-cyclically monotonicity, for a given cost function c, as follows.

Definition 3.2.8 ($c-$cyclical monotonicity)**.** We will say that a set $\Pi \subset \mathcal{X} \times \mathcal{Y}$ is $c-$cyclically monotone if, for any $(x_i, y_i) \in \Pi$, with $1 \leq i \leq N$, implies

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{\sigma(i)}), \qquad \text{for all permutations } \sigma \text{ of } \{1, \cdots, N\}$$

In the same fashion, we can define a kind of generalization of Fenchel-Legendre transforms, as it is presented in the following defintion.

Definition 3.2.9 (c-transforms)**.** Let $\phi : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Its c_+-transform, $\phi^{c+} : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ and its c_--transform, $\phi^{c-} : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ are defined as

$$\phi^{c+}(y) := \inf_{x \in \mathcal{X}} \big\{c(x,y) - \phi(x)\big\}. \quad \text{and} \quad \phi^{c-}(y) := \sup_{x \in \mathcal{X}} \big\{ - c(x,y) - \phi(x)\big\}.$$

respectively. In the same fashion, for any function $\psi : \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$, c_+-transform, $\psi^{c+} : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ and its c_--transform, $\psi^{c-} : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ are defined

$$\psi^{c+}(y) := \inf_{y \in \mathcal{Y}} \big\{c(x,y) - \psi(y)\big\}, \quad \text{and} \quad \psi^{c-}(y) := \sup_{y \in \mathcal{Y}} \big\{ - c(x,y) - \psi(y)\big\}.$$

Imitating the classical notions of Convex Analysis, we can define the properties of concavity and convexity with respect to a cost function c, based on definition of c-transforms as follows.

Definition 3.2.10 (c-concavity)**.** We will say that a function $\phi : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ is c-concave if there exists a function $\psi : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ such that $\phi = \psi^{c+}$. Similarly, we will say that a function $\psi : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ is c-concave it there exists a function $\phi : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ such that $\psi = \phi^{c+}$.

Definition 3.2.11 (c-convexity)**.** We will say that a function $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is c-convex if there exists a function $\psi : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ such that $\phi = \psi^{c-}$. Similarly, we will say that a function $\psi : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is c-convex it there exists a function $\phi : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ such that $\psi = \phi^{c-}$.

Remark 3.2.12 (Stability of c-concavity/convexity)**.** An interesting observations on above definitions is that "c-prodecures" are stable. This means that $\phi : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ is c-concave if and only if $\phi^{c+c+} = \phi$. The same result can be also obtained for $c-$convex functions.

Under the light of above generalizations of Convex Analysis notions, we are able to define the analogous notions of superdifferential and subdifferential, as the following definition witness.

Definition 3.2.13 (c-superdifferential and c-subdifferential)**.** Let $\phi : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ be a c-concave function. The c-superdifferential of ϕ, $\partial^{c+}\phi \subset \mathcal{X} \times \mathcal{Y}$ is defined as

$$\partial^{c+}\phi := \big\{(x,y) \in (\mathcal{X} \times \mathcal{Y}) : \phi(x) + \phi^{c+}(y) = c(x,y)\big\}.$$

In the same fashion, if $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a c-convex function is defined as

$$\partial^{c-}\phi := \big\{(x,y) \in \mathcal{X} \times \mathcal{Y} : \phi(x) + \phi^{c+}(y) = -c(x,y)\big\}.$$

An interesting and direct consequence of the above definition is that the c-superdifferential of a c-concave function is always a $c-$cyclically monotone set. Indeed, if $(x_i, y_i) \in \partial^{c+}\phi$ it holds that

$$\sum_{i=1}^{N} c(x_i, y_i) = \sum_{i=1}^{N} \phi(x_i) + \phi$$

Having this machinery in our toolbox, we able to prove several desired properties of transport plans. However, before doing this and in order to understand better the nature of these generalizations, let us remark a few things.

Remark 3.2.14 (Convex Analysis analogies)**.** As it was previously partial mentioned, the above definitions are nothing more but suitable generalizations of the basic definitions of Convex Analysis, in a more abstract setting with respect to the cost function c. Accurately, let us see how they can be reduced to classical definitions. Consider that $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and $c(x,y) = -\langle x, y\rangle$. Then one can observe that

1. a set is c-cyclically mononote, if and only if it is cyclically monotone
2. a function is c-convex (resp. c-concave) if and only if it is convex and lower semicontinuous (resp. concave and upper semicontinuous)
3. the c- subdifferential of a c-convex (resp. c-superdifferential of a c-concave) function is the classical subdifferential (resp. superdifferential)
4. the c_- is the classical Fenchel-Legendre transform, i.e. $\phi^*(y) = \sup_{x\in\mathcal{X}}\{-\langle x, y\rangle - \phi(x)\}$.

Remark 3.2.15 (Equivalent characterization of c-superdifferential)**.** A useful equivalent characterization of c-superdifferentiability is that a $y \in \mathcal{Y}$ lives in $\partial^{c+}\phi$ if and only if it holds that

$$\phi(x) = c(x,y) - \phi^{c+}(y) \text{ or equivalently } \phi(z) \leq (z,y) - \phi^{c+}(y), \text{ for all } z \in \mathcal{X}$$

or equivalently

$$\phi(x) - c(x,y) \geq \phi(z) - c(z,y), \quad \text{for every } z \in \mathcal{X}.$$

As a direct consequence, we have that the c-superdifferential of a c-concave function is always a c-cyclically monotone set. Indeed, if $(x_i, y_i) \in \partial^{c+}\phi$, it holds that

$$\sum_{i=1}^N c(x_i,y_i) = \sum_{i=1}^N \phi(x_i) + \phi^{c+}(y_i) = \sum_{i=1}^N \phi(x_i) + \phi^{c+}(y_i) = \sum_{i=1}^N \phi(x_i) + \phi^{c+}(y_{\sigma(i)} \leq \sum_{i=1}^N c(x_i, y_{\sigma(i)}),$$

for every permuations σ of $\{1, 2, \cdots, N\}$.

Now we are ready to prove the most important result of this section, that is the *fundamental theorem of Optimal Transport.* This result provide us a nice structural intuition about optimal transport plans and reveals many desired properties of them. Precisely, we will show that the optimality of a transport plan, depends only on the geometry of its support, and not on the distribution of its mass itself.

Theorem 3.2.16 (Fundamental Theorem of Optimal Transport)**.** *Consider two probability measures $\mu \in \mathscr{P}(\mathcal{X})$ and $\nu \in \mathscr{P}(\mathcal{Y})$ and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ continuous and bounded from below, such that*

$$c(x,y) \leq a(x) + b(y) \tag{3.10}$$

for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$. Also, let $\pi \in \Pi(\mu,\nu)$. Then the following are equivalent:

(i) the transportation plan π is optimal

(ii) the set $\mathrm{supp}(\pi)$ *is c-cyclically monotone*

(iii) there exists a c-concave function ϕ such that $\max\{\phi, 0\} \in L^1(\mu)$ *and* $\mathrm{supp}(\pi) \subset \partial^{c+}\phi$.

Proof. At first, one can observe that for any $\tilde{\pi} \in \Pi(\mu,\nu)$, the function $\max\{\phi, 0\}$ is integrable. Indeed, we have that

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\tilde{\pi}(x,y) \leq \int_{\mathcal{X}\times\mathcal{Y}} a(x) + b(x)d\tilde{\pi}(x,y) = \int_{\mathcal{X}} a(x)d\mu(x) + \int_{\mathcal{Y}} b(y)d\nu(y) < \infty, \tag{3.11}$$

for any $\tilde{\pi} \in \Pi(\mu, \nu)$. Now, combining 3.10 with 3.11, we can derive the integrability of $\max\{\phi, 0\}$. This fact together with the boundness from below of c, gives us that $c \in L^1(\pi)$ for any $\pi \in \Pi(\mu, \nu)$.

$(i) \Rightarrow (ii)$: To attack to this implication, we will follow the steps: for the sake of contradiction we will assume that the $\operatorname{supp} \pi$ is not c-cyclically monotone, we will construct a plan which will be positive and will have marginals μ, ν, and then we will prove that it is better that the optimal one: a contradiction.

So, following this strategy, let us assume that $\operatorname{supp} \pi$ is not c-cyclically monotone. In this case, we can find $N \in \mathbb{N}$, a family $\{(x_i, y_i)\}_{i=1}^N \subset \operatorname{supp} \pi$ and a permutation $\sigma \subset \{1, \cdots, N\}$ such that

$$\sum_{i=1}^N c(x_i, y_i) > \sum_{i}^N c(x_i, y_{\sigma(i)}).$$

At this point, we can take $\epsilon > 0$ such that

$$0 < \epsilon < \frac{1}{2N}\Bigg(\sum_{i=1}^N c(x_i, y_i) - \sum_{i=1}^N c(x_i, y_{\sigma(i)})\Bigg). \tag{3.12}$$

Thanks to continuity of c, there exists $r > 0$ such for every $1 \leq i \leq N$ and for every $(x, y) \in B_r(x_i) \times B_r(y_i)$ we have that

$$c(x, y) > (x_i, y_i) - \epsilon. \tag{3.13}$$

and for every $(x, y) \in B_r(x_i) \times B_r(y_{\sigma(i)})$ we have that

$$c(x, y) < c(x_i, y_{\sigma(i)}) + \epsilon. \tag{3.14}$$

Let us divide the rest of the proof into 3 steps, according to the strategy which has been already mentioned.

Step 1: *Construction of suitable measure.*

We consider the $V_i = B_r(x_i) \times B_r(y_i)$, which, by construction, consist an open set with respect metric topology of the product space $\mathcal{X} \times \mathcal{Y}$. Notice also that $\pi(V_i) > 0$, for every $i \in \mathbb{N}$, since we assume that $(x_i, y_i) \in supp(\pi)$.

Based on definition of V_i's, we define the measures

$$\pi_i = \frac{\pi|_{V_i}}{\pi(V_i)}$$

and

$$\mu_i = (\mathsf{p}_{\mathcal{X}})_{\#}\pi_i, \quad \nu_i = (\mathsf{p}_{\mathcal{Y}})_{\#}\pi_i$$

Taking $0 < a < \frac{1}{k}\min_{i \in \mathbb{N}} \pi(V_i)$, we construct a measure $\overline{\pi}_i \in \Pi(\mu_i, \nu_{\sigma(i)})$ such that $\overline{\pi}_i = \mu_i \otimes \nu_i$. Then we define a measure

$$\overline{\pi} := a\sum_{i=1}^k \pi_i + a\sum_{i=1}^k \overline{\pi}_i.$$

First, we have to prove that π is a transport plan, which in our case means that is positive and have marginals μ, ν respectively. Then we will get the contradiction if $\overline{\pi}$ is "more" optimal π (which by our assumption is optimal), or in other words

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

Step 2: *$\overline{\pi}$ is positive measure and has μ and ν as marginals*

Firstly, we claim that $\overline{\pi}$ is positive measure. Indeed, by construction of $\overline{\pi}$, we have only to check that $\pi - a\sum_{i=1}^{N}\pi_i$ is positive. To check this we have that

$$\pi - a\sum_{i=1}^{N}\pi_i > 0 \Rightarrow \pi - a\sum_{i=1}^{N}\frac{\pi|_{V_i}}{\pi(V_i)} > 0 \Rightarrow \pi - a\frac{k\pi|_{V_i}}{\pi(V_i)} > 0 \Rightarrow \pi - ak\pi_i \Rightarrow \frac{\pi}{k} < a\pi_i.$$

So, for positiveness, we have to prove only that $\frac{\pi}{N} < a\pi_i$. This inequality is true, since $k \in \mathbb{N}$, and based on definition of a and the definition of π_i we have

$$a\pi_i = \pi_{V_i} \quad \text{and} \quad \frac{a}{\pi_i} \leq \frac{1}{N},$$

So, combining the above inequalities, the desired positiveness follows.

Secondly, we claim that $\overline{p}$ has μ and ν as marginals. Indeed, by definition of $\overline{\pi}$ and using the definition of push-forward measure, we have

$$(\mathsf{p}_{\mathcal{X}})_{\#}\overline{\pi} = \mu - a\sum_{i=1}^{N}(\mathsf{p}_{\mathcal{X}})_{\#}\pi_i + \sum_{i=1}^{N}(\mathsf{p}_{\mathcal{X}})_{\#}\overline{\pi}_i = \mu - a\sum_{i=1}^{N}\mu_i + a\sum_{i=1}^{N}\mu_i = \mu$$

and in the same fashion

$$(\mathsf{p}_{\mathcal{Y}})_{\#}\overline{\pi} = \nu - a\sum_{i=1}^{k}(\mathsf{p}_{\mathcal{Y}})_{\#}\pi_i + \sum_{i=1}^{k}(\mathsf{p}_{\mathcal{Y}})_{\#}\overline{\pi}_i = \nu - a\sum_{i=1}^{k}\nu_i + a\sum_{i=1}^{k}\nu_i = \nu$$

Step 3: *$\overline{\pi}$ violates optimality*

Our final step is to prove that the constructed measure $\overline{\pi}$ violates optimality. To prove this statement, we will consider the difference

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi - \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\overline{\pi}$$

and prove that it is positive. So, using (3.13),(3.14), we have

$$\begin{aligned}
\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi - \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\overline{\pi} &= a\sum_{i=1}^{N}\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi_i - a\sum_{i=1}^{N}\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\overline{\pi}_i \\
&\geq a\sum_{i=1}^{N}(c(x_i,y_i)-\epsilon) - a\sum_{i=1}^{N}(c(x_i,y_{\sigma(i)})+\epsilon) \\
&= a\Big(c(x_i,y_i) - \sum_{i=1}^{N}(c(x_i,y_{\sigma(i)}) - 2N\epsilon\Big) \\
&> 0,
\end{aligned}$$

So $\overline{\pi}$ is "more" optimal π, which is optimal by our assumption. This is the condradiction, which completes our proof this implication.

$(ii) \Rightarrow (iii)$: For this implication, it suffices to prove that if $\Pi \subset \mathcal{X}\times\mathcal{Y}$ is a c-cyclically monotone set, then there exists a c-concave functtion ϕ such that $\Pi \subset \partial^{c+}\phi$ and $\max\{\phi, 0\} \in L^1(\mu)$. Let us fix $(\bar{x}, \bar{y}) \in \Pi$. One can observe that, since we want ϕ to be c-concave with its c-superdifferential containing Π, for any choice of $(x_i, y_i) \in \Pi$ with $1 \leq i \leq N$, we need to have

$$\begin{aligned}
\phi(x) &\leq c(x, y_1) - \phi^{c+}(y_1) = c(x, y_1) - c(x_1, y_1) + \phi(x_1) \\
&\leq \big(c(x, y_1 - c(x_1, y_1)\big) + c(x_1, y_2) - \phi^{c+}(y_2) \\
&= \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + \phi(x_2) \\
&\leq \cdots \\
&\leq \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2)\big) + \cdots + \big(c(x_N, \bar{y}) - c(\bar{x}, \bar{y})\big) + \phi(\bar{x}).
\end{aligned}$$

Based on these calculations, it is natural to define ϕ as the infimum of the above expression as $\{(x_i, y_i)\}_{1 \leq i \leq N}$ runs among all $N - ples$ in Π and N runs in $\mathbb{N}$. In addition, since we are free to add a constant to ϕ, we can forget about the term $\phi(\bar{x})$ and define

$$\phi(x) := \inf \Big\{ \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2 - c(x_2, y_2)\big) + \cdots + \big(c(x_N, \bar{y}) + c(\bar{x}, \bar{y})\big) \Big\},$$

where the infimum is taken on $(x_i, y_i) \in \Pi$, for $1 \leq i \leq N$, and $N \leq 1$. Choosing now $N = 1$ and $(x_1, y_1) = (\bar{x}, \bar{y})$, we obtain that $\phi(\bar{x}) \leq 0$. In reverse direction, due to c-cyclical monotonicity of Π, we obtain that $\phi(\bar{x}) \geq 0$. So finally, $\phi(\bar{x}) = 0$.

Now, by its definition, it is clear that ϕ is c- concave. Employing the same trick as before by choosing $N = 1$ and $(x_1, y_1) = (\bar{x}, \bar{y})$, and using the alternative characterization of superdifferential given by Remark 3.2.15, we obtain that

$$\phi(x) \leq c(x, \bar{y}) - c(\bar{x}, \bar{y}) \leq a(x) + b(\bar{y}) - c(\bar{x}, \bar{y}).$$

This fact, combining with $a \in L^1(\mu)$, yields that $\max\{\phi, 0\} \in L^1(\mu)$. It left only to prove that $\partial^{c+}\phi$ contains Π. For see this, we choose $(\tilde{x}, \tilde{y}) \in \Pi$, and consider that $(x_1, x_2) = (\tilde{x}, \tilde{y})$. Again, by definition, of ϕ, we have

$$\begin{aligned}
\phi(x) &\leq c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + \inf \Big\{ \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + \cdots + \big(c(x_N, \bar{y}) + c(\bar{x}, \bar{y})\big) \Big\} \\
&= c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + \phi(\tilde{x}).
\end{aligned}$$

Again, by Remark 3.2.15, this inequality shows that $(\tilde{x}, \tilde{y}) \in \partial^{c+}\phi$, as our goal was.

(iii) $\Rightarrow$ *(i)* : To prove this implication, we consider that $\tilde{\pi} \in \Pi(\mu, \nu)$, a is a transport plan. We want to prove that

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

We recall that

$$\phi(x) + \phi^{c+}(y) = c(x, y) \text{ for any } (x, y) \in \operatorname{supp}(\pi) \quad \text{and} \quad \phi(x) + \phi^{c+}(y) \leq c(x, y), \text{ for } x \in \mathcal{X}, y \in \mathcal{Y}.$$

Therefore, we have

$$\begin{aligned}
\int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) &= \int_{\mathcal{X} \times \mathcal{Y}} \phi(x) + \phi^{c+}(y) d\pi(x, y) \\
&= \int_{\mathcal{X}} \phi(x) d\mu(x) + \int_{\mathcal{Y}} \phi^{c+}(y) d\nu(y) \\
&= \int_{\mathcal{X} \times \mathcal{Y}} \phi(x) + \phi^{c+} d\tilde{\pi}(x, y) \\
&\leq \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\tilde{\pi}(x, y),
\end{aligned}$$

which makes our proof complete. □

3.3 Duality Framework

Many problems in optimization are closely related with a different, but complementary problem, their *dual* problem. The duality relation between the primal and the dual problem could be a powerful tool in order to study these problems. The same fact holds also with respect to optimal transportation problem. Usually, as it widely known, the dual objects of functions are measures. In optimal transportation problems, perhaps, due to historically reasons, the primal dual consists essentially an optimization problem over probability measures. So, intuitively, it will be an optimization problem over functionals. To make the discussion a little bit more technical, let us define the set

$$\Phi_c = \big\{(\phi, \psi) \in L^1(\mu) \times L^1(\nu) : \phi(x) + \psi(y) \leq c(x, y)\big\} \tag{3.15}$$

in μ-a.e. sense for every $x \in \mathcal{X}$ and $\nu-$a.e. sence for every $y \in \mathcal{Y}$. Note that, in many situations, we condition $\phi(x) + \psi(y) \leq c(x, y)$ is called *cost condition,* due to its economic implementation. Under this perspective, we introduce functionals

$$\mathcal{I}(\pi) = \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) d\pi(x, y) \tag{3.16}$$

and based on (3.15), we introduce the functional

$$\mathcal{J}(\phi, \psi) = \int_{\mathcal{X}} \phi(x) d\mu(x) + \int_{\mathcal{Y}} \psi(y) d\nu(y) \tag{3.17}$$

Before of the closer examination of duality relations in optimal transport problem, we recall a famous result of Convex Analysis, in the following proposition.

Theorem 3.3.1 (Fenchel-Rockafellar). *Let V be a normed space and V^* its (topological) dual. Consider also two convex functions $\Phi, \Psi : V \to \mathbb{R} \cup \{+\infty\}$ and Φ^*, Ψ^* their Legendre transforms respectively. We assume that there exists $z_0 \in V$ such that $\Phi(z_0) < \infty$ and $\Psi(z_0) < \infty$ and Φ is continuous at z_0. Then it holds that*

$$\inf_{z\in V} \Big\{\Phi(x) + \Psi(x)\Big\} = \max_{z^*\in V*} \Big\{ - \Phi^*(z^*) - \Psi^*(z^*)\Big\} \tag{3.18}$$

Proof. We observe that it suffiecient to prove that

$$\sup_{z^*\in V^*} \inf_{x,y\in V} \Big\{\Phi(x) + \Psi(y) + \langle z^*, x - y\rangle\Big\} = \inf_{x\in V} \Big\{\Phi(x) + \Psi(y)\Big\}. \tag{3.19}$$

If we choose $x = y$, we can see that the left hand quantity cannot be larger that the right hand. Based on this fact we have just to proof the existence of a linear form $z^* \in V$ such that

$$\Phi(x) + \Psi(y) + \langle z^*, x - y\rangle \geq \inf_{x\in V} \Big\{\Phi(x) + \Psi(y)\Big\} = m. \tag{3.20}$$

Since there exists a $z_0 \in V$ such that $\Phi(z_0) < \infty$ and $\Psi(z_0) < \infty$, we observe that infimum m is finite.

Now, let us define

$$C = \big\{(x, \lambda) \in V \times \mathbb{R} : \lambda > \Phi(x)\big\} \quad \text{and } C' = \big\{(y, \mu) \in V \times \mathbb{R} : \mu \leq m - \Psi(y)\big\}. \tag{3.21}$$

Obviously, C, C' are convex sets, since Φ, Ψ are convex functions. From our assumptions, we can deduce that $(z_0, \Phi(z_0) + 1) \in \operatorname{Int} C \neq \varnothing$. So, since C has not empty interior, this implies that

$\overline{C} = \overline{\operatorname{Int} C}$. Moreover, because of the construction of the infimum m, C and C' are disjoint. Thanks to Hahn-Banach theorem, there exists a (nontrivial) linear form $h \in (V \times \mathbb{R})^*$ such that

$$\inf_{c\in C} \langle h, c\rangle = \inf_{c\in \operatorname{Int} C} \langle h, c\rangle \geq \sup_{c'\in C'} \langle h, c'\rangle. \tag{3.22}$$

Equivalently, there exists $w^* \in V^*$ and $\alpha \in \mathbb{R}$, with $(w^*, \alpha) \neq (0,0)$, such that

$$\langle w^*, x\rangle + \alpha\lambda \geq \langle w^*, y\rangle + \alpha\mu$$

as soon as $\lambda > \Phi(x)$ and $\mu \leq m - \Psi(y)$. This is only possible only when $\alpha > 0$, and thus, with $z^* = w^*/\alpha$ we have

$$\langle z^*, z\rangle + \lambda \leq \langle z^*, y\rangle + \mu$$

or equivalently,

$$\langle z^*, x\rangle + \Phi x \geq \langle z^*, y\rangle + m - \Psi(y).$$

Since this holds for all $x, y \in V$, we conclude that (5.23) is true and our proof is complete. $\square$

Now we can state and prove the well-known *Kantorovich duality theorem.*

Theorem 3.3.2 (Kantorovich duality)**.** *Consider that $\mathcal{X}$ and $\mathcal{Y}$ be complete and separable metric spaces, $\mu \in \mathscr{P}(\mathcal{X})$ and $\nu \in \mathscr{P}(\mathcal{Y})$ is two probability measures and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a lower-semicontinuous cost function. Then*

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}(\gamma) = \sup_{(\phi,\psi)\in\Phi^c} \mathcal{J}(\phi, \psi). \tag{3.23}$$

Furthermore, the infimum is a minimum.

Proof. Following the spirit of Villani (2003), we divide the proof into 3 steps.

Step 1: *Compact case.*

At first, we assume that $\mathcal{X}, \mathcal{Y}$ are compact spaces and c is continuous. Let $V = C_b(\mathcal{X} \times \mathcal{Y})$, the set of all continuous functions equipped with supremum norm $\|\cdot\|_\infty$. By Riesz representation theorem, its (topological) dual can be isomorphically identified by the set of all (regular) Radon measures $V^* = \mathscr{M}(\mathcal{X} \times \mathcal{Y})$ normed by total variation $\|\cdot\|_{TV}$. We define the maps Φ, Ψ on $C_b(\mathcal{X} \times \mathcal{Y})$ by

$$\Phi(u) = \begin{cases} 0, & \text{if } u(x,y) \geq -c(x,y) \\ +\infty, & \text{else} \end{cases} \quad \text{and } \Psi(u) = \begin{cases} \displaystyle\int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu, & \text{if } u(x,y) = \phi(x) + \psi(y) \\ +\infty & \text{else.} \end{cases}$$

Now, one can observe that the assumptions of Fenchel-Rockafellar theorem are satisfied with $z_0 = 1$, and so we obtain

$$\inf_{x,y\in V} \Big\{ \Phi(x) + \Psi(y) \Big\} = \max_{z^*\in V*} \Big\{ - \Phi^*(-z^*) - \Psi(z^*) \Big\}. \tag{3.24}$$

Let us compute both sides of (3.24). The left-hand-side we have that

$$\inf \Big\{ \int_{\mathcal{X}} \phi d\mu + \int_Y d\nu : \phi(x) + \psi(y) \geq -c(x,y) \Big\} = -\sup \big\{ \mathcal{J}(\phi, \psi) : (\phi, \psi) \in \Phi_c \big\}. \tag{3.25}$$

Computing Legendre-Fenchel tranform of Φ, for a given $\pi \in \mathscr{M}(\mathcal{X} \times \mathcal{Y})$, we obtain

$$\begin{aligned}\Phi^*(\pi) &= \sup_{u \in C_b(\mathcal{X}\times\mathcal{Y})} \Big\{ -\int_{\mathcal{X}\times\mathcal{Y}} u(x,y)d\pi(x,y) : u(x,y) \geq -c(x,y) \Big\} \\ &= \sup_{u \in C_b(\mathcal{X}\times\mathcal{Y})} \Big\{ \int_{\mathcal{X}\times\mathcal{Y}} \Big\{ \int u(x,y)d\pi(x,y) : u(x,y) \leq c(x,y) \Big\}.\end{aligned}$$

If π is not nonnegative measure then there exists a nonpositive function $v \in C_b(\mathcal{X} \times \mathcal{Y})$ such that $\int_{\mathcal{X}\times\mathcal{Y}} v d\pi(x,y) > 0$. Choosing $u = \lambda v$ and taking $\lambda \longrightarrow \infty$ we obtain that supremum is equal to $+\infty$. In different circumstances, if π is nonnegative, then the supremum is clearly equal to $\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y)$. Summarizing the above discussion Φ^* is equal to

$$\Phi^*(-\pi) = \begin{cases} \sup_{u \in C_b(\mathcal{X}\times\mathcal{Y})} \Big\{ \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y) \Big\} & \text{if } \pi \in \mathscr{M}_+(\mathcal{X} \times \mathcal{Y}) \\ +\infty, \quad \text{else} \end{cases}$$

and in the same fashion, Ψ^* is equal to

$$\Psi^*(\pi) = \begin{cases} 0, & \text{if } \int_{\mathcal{X}\times\mathcal{Y}} \phi(x) + \psi(y) d\pi(x,y) = \int_{\mathcal{X}} \phi d\mu - \int_{\mathcal{Y}} \psi d\nu, \quad \forall (\phi,\psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y}) \\ +\infty, \quad \text{else} \end{cases}$$

Changing signs and putting all together, we get

$$\inf_{\pi \in \Pi(\mu_1,\mu_2)} \mathcal{I}(\pi) = \sup_{(\phi,\psi) \in \Phi_c \cap C_b} \mathcal{J}(\phi,\psi)$$

Step 2: *Reduction to compact case*

We assume now that c is bounded and uniformly continuous. We will attack, at first, to "less or equal" inequality which is less technical and more more easier. We claim that.

$$\sup_{\Phi_c \cup C_b} \mathcal{J}(\phi,\psi) \leq \sup_{\Phi_c \cup L^1} \mathcal{J}(\phi,\psi) \leq \inf_{\pi \in \Pi(\mu,\nu)} \mathcal{I}(\pi).$$

To prove this claim, at first, we can the left-hand side inequality is trivial, since $C_b(\mathcal{X}) \times C_b(\mathcal{Y}) \subseteq L^1(\mu) \times L^1(\nu)$. For the right-hand inequality, let $(\phi,\psi) \in \Phi_c \subseteq L^1$ and $\pi \in \Pi(\mu,\nu)$. Then, by definition of π, we have that

$$\mathcal{J}(\phi,\psi) \leq \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu = \int_{\mathcal{X}\times\mathcal{Y}} \big(\phi(x) + \psi(y)\big) d\pi(x,y).$$

But ϕ, ψ satisfying the price condition $\phi(x) + \psi(y) \leq c(x,y)$ $\pi-$a.e. Indeed, let negligible sets N_x, N_y such that $\mu(N_x) = \nu(N_y) = 0$ and price condition holds for $(x,y) \in N_x^c \times N_y^c$. Since the probability measure π has marginals μ and ν, we can say that $\pi(N_x \times \mathcal{Y}) = \mu(N_x) = 0$ and $\pi(N_y \times \mathcal{X}) = \nu(N_y) = 0$ and hence $\pi\big((N_x^c \times N_y^c)^c\big) = 0$. So from this, it follows that

$$\int_{\mathcal{X}\times\mathcal{Y}} \big(\phi(x) + \psi(y)\big) d\pi(x,y) \leq \int_{\mathcal{X}\times\mathcal{Y}} c(x,y) d\pi(x,y) = \mathcal{I}(\pi).$$

Now, taking supremum and infimum in left-hand side and right-hand side respectively, we obtain that

$$\sup_{\Phi_c \cap L^1} \mathcal{J}(\phi,\psi) \leq \inf_{\pi \in \Pi(\mu,\nu)} \mathcal{I}(\pi) \tag{3.26}$$

which proves the "less or equal" inequality.

Let us now attack to the reverse inequality, which is a bit more technical and tricky. We define $\|c\|_\infty = \sup_{\mathcal{X}\times\mathcal{Y}} c(x,y)$. We will reduce this case to the compact one by a careful truncation procedure. Consider that π^* is the optimal transport plan, i.e.

$$\mathcal{I}(\pi^*) = \inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}(\pi).$$

where the above infimum is finite, since c is bounded. The existence of such a π^* is guaranteed by compactness of $\Pi(\mu,\nu)$, based on Theorem 3.2.4 (Prokhorov' theorem), where details can be found in step 3. Consider now $\delta > 0$. Since $\mathcal{X}\times\mathcal{Y}$ are complete and separable metric spaces, the product space, since it is a finite product, is also complete and separable. In particular, π^* is tight and there exist compact sets $K_o \subseteq \mathcal{X}$ and $F_o \subseteq \mathcal{Y}$ such that $\mu(K_o^c) \le \delta$ and $\nu(F_o^c) \le \delta$, so we have

$$\pi^*\big((K_o\times F_o)^c\big) \le 2\delta.$$

We define a probability measure π_o^* on $K_o\times F_o$ by

$$\pi_o^* = \frac{\mathbb{1}_{K_o\times F_o}}{\pi^*(K_o\times F_o)}\pi^*$$

If μ_o, ν_o are the marginals of π_o^* onto $K_o\times F_o$ respectively, consider the set $\Pi_o(\mu_o,\nu_o)$, i.e. the set of probability measure on $K_o\times F_o$ having μ_o,ν_o as marginals. We define

$$\mathcal{I}_o(\pi_o) := \int_{K_o\times F_o} c(x,y)d\pi_o(x,y).$$

Consider now $\tilde{\pi_o} \in \Pi_o(\mu_o,\nu_o)$ such that

$$\mathcal{I}_o(\tilde{\pi_o} = \inf_{\pi_o\in\Pi_o(\mu_o,\nu_o)} \mathcal{I}_o(\pi_o).$$

We construct a $\tilde{\pi} \in \Pi(\mu,\nu)$ from $\tilde{\pi_o}$ by

$$\tilde{\pi} := \pi^*(K_o\times F_o)\cdot\tilde{\pi_o} + \mathbb{1}_{(K_o\times F_o)^c}\cdot\pi^*.$$

So, based on this fact we obtain

$$\begin{aligned}\mathcal{I}(\tilde{\pi}) &= \pi^*(K_o\times F_o)\mathcal{I}(\tilde{\pi_o} + \int_{(K_o\times F_o)^c} c(x,y)d\pi^*(x,y)\\ &\le \mathcal{I}_o(\tilde{\pi_o}) + 2\delta\|c\|_\infty\\ &\le \inf_{\pi_o\in\Pi_o(\mu_o,\nu_o)} \mathcal{I}_o(\pi_o) + 2\delta\|c\|_\infty,\end{aligned}$$

and so it follows that

$$\inf_{\pi\in\Pi(\mu,\nu)}\mathcal{I}(\pi) \le \inf_{\pi_o\in\Pi_o(\mu_o,\nu_o)}\mathcal{I}_o(\pi_o) + 2\delta\|c\|_\infty.$$

We introduce now the functional $\mathcal{J}_o$ on $L^1(\mu_o)\times L^1(\nu_o)$ defined as

$$\mathcal{J}_o(\phi_o,\psi_o) := \int_{K_o}\phi_o d\mu_o + \int_{F_o}\psi_o d\nu_o.$$

By step 1, we have that $\inf\mathcal{I}_o = \sup J_o$, where supremum runs over all admissible functions $(\phi_o,\psi_o) \in L^1(\mu_o)\times L^1(\nu_o)$, satisfying the price condition i.e. $\phi_o(x)+\psi_o(y) \le c(x,y)$ a.e. for all x,y. In particular, for $\delta>0$ there exists a couple of admissible functions $\tilde{\phi}_o, \tilde{\psi}_o$ such that

$$\mathcal{J}_o(\tilde{\phi}_o,\tilde{\psi}_o) \ge \sup\mathcal{J}_o - \delta.$$

Our strategy is the following: we will construct an efficient couple (ϕ, ψ), from $(\tilde{\phi}_o, \tilde{\psi}_o)$, for the maximization of $\mathcal{J}(\phi, \psi)$. We can ensure that price condition $\tilde{\phi}_o + \tilde{\psi}_o \leq c(x, y)$ holds, not only in a.e. set, but for all x, y, provided that $\tilde{\phi}_o, \tilde{\psi}_o$ can take values in $\mathbb{R} \cup \{-\infty\}$. To see that, consider N_x, N_y are two negligible sets such tat the price condition holds for $(x, y) \in N_x^c \times N_y^c$. Then, if we redefine the values of ϕ, ψ to be $-\infty$ on N_x and N_y respectively, the notion of a.e. walks away.

Without loss of generality, let us assume that $\delta \leq 1$. Then, since $\mathcal{J}_o(0, 0) = 0$, we can see that $\sup \mathcal{J}_o = 0$ and hence $\mathcal{J}_o(\tilde{\phi}_o, \tilde{\psi}_o) \geq \delta \geq 1$. By writing

$$\mathcal{J}_o(\tilde{\phi}_o, \tilde{\psi}_o) = \int_{\mathcal{X} \times \mathcal{Y}} \big(\tilde{\phi}_o(x) + \tilde{\psi}_o(y)\big) d\pi_o(x, y)$$

where $\pi_o \in \Pi_o(\mu_o, \nu_o)$, we deduce that there exists $(x_o, y_o) \in K_o \times F_o$ such that $\tilde{\phi}_o(x_o) + \tilde{\psi}_o(y_o) \geq -1$. Then if we replace $(\tilde{\phi}_o, \tilde{\psi}_o)$ by $(\tilde{\phi}_o + s, \tilde{\psi}_o - s)$ for some $s \in \mathbb{R}$, we do not change the value of functional $\mathcal{J}_o(\tilde{\phi}_o, \tilde{\psi}_o)$ and the resulting couple is still admissible. By a proper choice of s, we can obtain

$$\tilde{\phi}_o(x_o) \geq \frac{1}{2} \quad \text{and} \quad \tilde{\psi}_o(x_o) \geq -\frac{1}{2}.$$

This implies that for every $(x, y) \in K_o \times F_o$ we have that

$$\begin{aligned} \tilde{\phi}_o(x) &\leq c(x, y_o) - \tilde{\psi}_o(x_o) \leq c(x, y_o) + \frac{1}{2} \\ \tilde{\psi}_o(y) &\leq c(x_o, y) - \tilde{\phi}_o(x_o) \leq c(x_o, y) + \frac{1}{2}. \end{aligned}$$

Now we have to "improve", in some sense, admissible couples. The following key trick is due Rüschendorf. Let us define for a $x \in \mathcal{X}$,

$$\overline{\phi}_o(x) := \inf_{y \in F_o} \big\{c(x, y) - \tilde{\psi}_o(y)\big\}$$

We can easily see that $\tilde{\phi}_o \leq \overline{\phi}_o$ on $\mathcal{X}_o$. This implies that $\mathcal{J}_o(\overline{\phi}_o, \tilde{\psi}_o) \geq \mathcal{J}_o(\tilde{\phi}_o, \tilde{\psi}_o)$. Moreover, for all $x \in \mathcal{X}$ we control $\overline{\phi}_o(x)$ both from above and below, that is

$$\begin{aligned} \overline{\phi}_o &\geq \inf_{y \in F_o} \big\{c(x, y) - c(x_o, y)\big\} - \frac{1}{2} \\ \overline{\phi}_o &\leq c(x, y_o) - \tilde{\psi}_o(y_o) \leq c(x, y_o) + \frac{1}{2} \end{aligned}$$

Similarly, we define, for a $y \in \mathcal{Y}$

$$\overline{\psi}_o(y) := \inf_{x \in K_o} \big\{c(x, y) - \overline{\phi_o}(x)\big\},$$

and still having that $(\overline{\phi}_o, \overline{\psi}_o) \in \Phi_c$. Then, we infer

$$\mathcal{J}_o(\overline{\phi}_o, \overline{\psi}_o) \geq \mathcal{J}_o(\overline{\phi}_o, \psi_o) \geq \mathcal{J}_o(\tilde{\phi}_o, \tilde{\psi}_o).$$

Similarly with $\overline{\phi}_o$, we can control $\overline{\psi}_o$ both from above and below, that is, for all $y \in \mathcal{Y}$

$$\begin{aligned} \overline{\psi}_o &\geq \inf_{x \in K_o} \big\{c(x, y) - c(x, y_o)\big\} - \frac{1}{2} \\ \overline{\psi}_o &\leq c(x_o, y) - \overline{\phi}_o(y_o) \leq c(x_o, y) - \tilde{\phi}_o(y_o \leq c(x_o, y) + \frac{1}{2}. \end{aligned}$$

Moreover, we obtain that

$$\overline{\phi}_o(x) \geq -\|c\|_\infty - \frac{1}{2} \quad \text{and} \quad \overline{\psi}_o(y) \geq -\|c\|_\infty - \frac{1}{2}.$$

Now, combining all of these bounds we have

$$\begin{aligned}
\mathcal{J}(\overline{\phi}_o, \overline{\psi}_o) &= \int_{\mathcal{X}} \overline{\phi}_o d\mu + \int_{\mathcal{Y}} \overline{\psi}_o d\nu = \int_{\mathcal{X}\times\mathcal{Y}} \big(\overline{\phi}_o(x) + \overline{\psi}_o(y)\big) d\pi^*(x,y) \\
&= \pi^*(K_o \times F_o) \int_{K_o \times F_o} \big(\overline{\phi}_o(x) + \overline{\psi}_o(y)\big) d\pi^*_o(x,y) + \int_{(K_o\times F_o)^c} \big(\overline{\phi}_o(x) + \overline{\psi}_o(y)\big) d\pi^*(x,y) \\
&\geq (1-2\delta)\bigg(\int_{K_o} \overline{\phi}_o d\mu_o + \int_{F_o} \overline{\psi}_o d\nu_o\bigg)(2\|c\|_\infty)\pi^*\big((K_o \times F_o)^c\big) \\
&\geq (1-2\delta)\mathcal{J}_o(\overline{\phi}_o, \overline{\psi}_o) - 2\delta(2\|c\|_\infty + 1) \\
&\geq (1-2\delta)\mathcal{J}_o(\tilde{\phi}_o, \tilde{\psi}_o) - 2\delta(2\|c\|_\infty + 1) \\
&\geq (1-2\delta)(\inf \mathcal{I} - \delta) - 2\delta(2\|c\|_\infty + 1) \\
&\geq (1-2\delta)\big(\inf \mathcal{I} - (2\|c\|_\infty + 1)\delta\big) - 2\delta(2\|c\|_\infty + 1).
\end{aligned}$$

Since $\delta \leq 1$ is arbitary small, we conclude, taking account the inequality in (3.26), we conclude that $\sup \mathcal{J} = \inf \mathcal{I}$, which completes step 3.

Step 3: *General case*

Let us now define $c := \sup_n c_n$, where $(c_n)_{n\in\mathbb{N}}$ stands for a nondecreasing bounded sequence of nonnegative, uniformly continuous cost functions. Consider also the functional $\mathcal{I}_n$, defined on $\Pi(\mu,\nu)$ defined as

$$I_n(\pi) := \int_{\mathcal{X}\times\mathcal{Y}} c_n(x,y) d\pi(x,y).$$

By step 2, we have that

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}_n = \sup_{(\phi,\psi)\in\Phi_{c_n}} \mathcal{J}(\phi,\psi). \tag{3.27}$$

We will show that

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}(\pi) = \sup_n \inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}_n(\pi) \tag{3.28}$$

and for every $n \in \mathbb{N}$,

$$\sup_{(\phi,\psi)\in\Phi_{c_n}} \mathcal{J}(\phi,\psi) \leq \sup_{(\phi,\psi)\in\Phi_c} \mathcal{J}(\phi,\psi). \tag{3.29}$$

Indeed, if we combine (3.27), (3.28) and (3.29), we obtain

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}(\pi) \leq \sup_{(\phi,\psi)\in\Phi_c} \mathcal{J}(\phi,\psi)$$

and since we know by 3.26 that the converse inequality is true, we get the equality. By construction, we have that $c_n \leq c$, obviously it follows that Φ_{c_n} is a subset of Φ_c, on which $\mathcal{J}_n$ coincides with $\mathcal{J}$, so (3.29) becomes trivial.

In consideration of construction of $\mathcal{I}_n$, as a nondecreasing sequence of functionals, it is clear that $(\inf \mathcal{I}_n)_{n\in\mathbb{N}}$ is a nondecreasing sequence, bounded from above by $\inf \mathcal{I}$. Therefore, we only have to prove that

$$\lim_{n\to\infty} \inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}_n(\pi) \leq \inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}(\pi). \tag{3.30}$$

As it has been proved in Proposition 3.2.5, the set of all transport plans $\Pi(\mu,\nu)$ is tight. Then, by Proposition 3.2.4 (Prokhorov's theorem), this implies that $\Pi(\mu,\nu)$ is relatively compact for the weak topology of probability measures. In particular, if $(\pi_{n_k})_{k\in\mathbb{N}}$ is a mimimizing sequence for $\inf \mathcal{I}_n(\pi)$, extracting a subsequence and keeping the same notation, π_{n_k} converges narrowly to some probability measure $\pi_n \in \mathscr{P}(\mathcal{X}\times\mathcal{Y})$ as $k \longrightarrow \infty$, i.e. there exists a $g \in C_b(\mathcal{X}\times\mathcal{Y})$ such that

$$\int_{\mathcal{X}\times\mathcal{Y}} g(x,y)d\pi_{n_k}(x,y) \overset{k\to\infty}{\longrightarrow} \int_{\mathcal{X}\times\mathcal{Y}} g(x,y)d\pi_n(x,y).$$

From that, we can see that $\pi_n \in \Pi(\mu,\nu)$ and

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}_n(\pi) = \lim_{n\to\infty}\int_{\mathcal{X}\times\mathcal{Y}} c_n(x,y)d\pi_{n_k}(x,y) = \int_{\mathcal{X}\times\mathcal{Y}} c_n(x,y)d\pi_n(x,y),$$

from which we obtain the existence of minimizing probability measure π_n. In this spirit, due compactness of $\Pi(\mu,\nu)$, the sequence $(\pi_n)_{n\in\mathbb{N}}$ admits a cluster point π^*. Whenever $n \geq m$, we have

$$\mathcal{I}_n(\pi_n) \geq \mathcal{I}_m(\pi_n),$$

and by continuity of $\mathcal{I}_m$, we have that

$$\lim_{n\to\infty} \mathcal{I}_n(\pi_n) \geq \limsup_{n\to\infty} \mathcal{I}_m(\pi_n) \geq I_m(\pi^*).$$

At last, by Monotone Convergence theorem, we have that $\mathcal{I}_m(\pi^*) \longrightarrow \mathcal{I}(\pi^*)$ as $m \to \infty$, so we obtain

$$\lim_{n\to\infty} \mathcal{I}_n(\pi_n) \geq \lim_{m\to\infty} \mathcal{I}_m(\pi^*) = \mathcal{I}(\pi^*) \geq \inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}(\pi),$$

which proves (3.30).

To conclude our proof, we will prove that infinum is really attained. This is again a collorary of compactness of $\Pi(\mu,\nu)$. Indeed, if $(\pi_k)_{k\in\mathbb{N}}$ is a minimizing sequence for $\mathcal{I}$ and π^* is a weak cluster point of $(\pi_k)_{k\in\mathbb{N}}$, employing Monotone Convergence theorem for the sequence $(c_n)_{n\in\mathbb{N}}$, we obtain

$$\mathcal{I}(\pi^*) = \lim_{n\to\infty} \mathcal{I}_n(\pi^*) \leq \lim_{n\to\infty}\lim_{k\to\infty} \mathcal{I}_n(\pi_k) \leq \limsup_{k\to\infty} \mathcal{I}(\pi_k) = \inf \mathcal{I},$$

which makes our proof complete. □

The duality relation of optimal transport problem has a plethora of many important and useful consequences. Between them, there are two which are related with c-concavity of the corresponding functional and a specific cost function induced by a metric on the base space. Both of them, because of their importance in what follows are presented in the following remarks.

Remark 3.3.3 (Duality and c-concavity)**.** An interesting property of the maximizing couple (ϕ,ψ) is dual problem is that in the form (ϕ,ϕ^{c^+}). To see this, the only one that we have to do is employing Theorem 3.2.16. To be more accurate, thanks to the later theorem, we have that there exist a c-concave function ϕ, susch that $\mathrm{supp}(\pi) \subset \partial^{c_+}\phi$, $\max\{\phi,0\} \in L^1(\mu)$ and $\max\{\phi^{c_+},0\} \in L^1(\nu)$. Under this consideration, and recalling the strategy of proof that this fact means optimality of $\pi \in \Pi(\mu,\nu)$, we have

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

Therefore, we have that $\phi \in L^1(\pi)$ and $\phi^{c_+} \in L^1(\nu)$, which means that the couple (ϕ,ϕ^{c_+}) is an admissible couple of the dual problem. Such as c-concave functions, which hold the property that the couple (ϕ, ϕ^{c_+}) is a maximizing couple for the dual problem are usually called *Kantorovich potentials.* Such kind of potentials enjoy two related but different imporant properties with respect to optimal transport problem. Precisely, at first, as we mentioned, they are maximizing functions in the dual problem, and secondly, the support of optimal transport plans, as we have seen in Theorem 3.2.16, is contained in the superdifferential of c-concave functions.

Remark 3.3.4 (Kantorovich-Rubinstein duality)**.** In case where $c = d$ is a distance, then a c^+-convex function is just a 1-Lipschitz function and its own c^+ transform. To see this, consider that if ψ is c^+ convex, then it is obviously 1-Lipschitz. On contrary, if ψ is 1-Lipschitz, then we have that $\psi(x) \leq \psi(y) + d(x,y)$, so we arrive the fact that $\psi(x) = \inf_y\{\psi(y) + d(x,y)\} = \psi^{c^+}(x)$. Under this pespective and recalling Theorem 3.3.2, we can lead to a important duality formula, that is

$$\mathcal{W}_1(\mu,\nu) = \sup_{\|\psi\|_{\mathrm{Lip}\leq 1}} \left\{ \int_{\mathcal{X}} \psi d\mu - \int_{\mathcal{X}} \psi d\nu \right\},$$

which is the so-called *Kantorovich-Rubinstein duality.*

3.4 Existence and Characterizations of Optimal Transport Maps

As we mentioned, Monge's relaxation of transportation problem from transport plans to transport maps thanks to its nature, allow us to use powerful tools from Variational and Convex Analysis. Under specific (and weak) assumptions, all of these chique tools make us able to obtain existence and uniqueness of solution of optimal transportation problem, and moreover, to understand several structural properties of its solution, which were strongly connected with its the topological and geometrical nature and provide us chique, abstract and quite general results.

Nevertheless, going backwards to looking for optimal transport maps, instead of plans, we have seen that in general this problem admits several degenerate cases (i.e. maybe there is no optimal transport map or every transport map is optimal). Having, this observation in our mind, a natural question arising is: under which additional assumptions can we find at least optimal transport plan which induced by an optimal transport map? In the other words, this question can be translated in terms of fine properties of the transport maps.

The first step of answering such a question relies on their interplay with the structure of optimal transport plans. Moreover, having as starting point the equivalence of optimality of a plan and the c-monotonicity of its support (Proposition 3.2.16), one can ask how "far", in some sense, is this support from being a graph of a transport map. Such a vague and unclear concern stands in the heart of the necessary assumptions which we mentioned above. This answer of this concern is relies on the following proposition.

Proposition 3.4.1 (Knott-Smith Optimality criterion)**.** *Consider $\mu, \nu \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ and $\pi \in \Pi(\mu,\nu)$ be an optimal transport plan. Then π induced by a measurable map $\boldsymbol{T} : \mathcal{X} \to \mathcal{Y}$ if and only if there exists a π-measurable set $\Gamma \subseteq \mathcal{X} \times \mathcal{Y}$ where π is concentrated, such that for μ-a.e. $x \in \mathcal{X}$ there exists only one $y = \boldsymbol{T}(x)$ with $(x,y) \in \Gamma$. In this case, π is induced by the map $\boldsymbol{T}$.*

Proof. We shall prove only the one implication.
$(\Leftarrow)$: Consider a π-measurable set $\Gamma \subseteq \mathcal{X} \times \mathcal{Y}$ where π is concentrated, such that for μ-a.e. $x \in \mathcal{X}$, $\boldsymbol{T} : \mathcal{X} \to \mathcal{Y}$ $N \subset \mathcal{X}$ be a μ-null set. Removing from Γ the product $N \times \mathcal{Y}$ we can assume that Γ consists a graph. Thanks $\mathcal{X}, \mathcal{Y}$ are complete and separable metric spaces, we have that μ, ν are automatically regular. Then, thanks to inner regularity of them, we can write Γ as

$$\Gamma = \bigcup_{n \in \mathbb{N}} \Gamma_n$$

which is σ-compact. Under this consideration, the domain of $\boldsymbol{T}$ is also σ- compact, hence it is a Borel set. Hence the restriction of $\boldsymbol{T}$ to the set $\mathsf{p}_{\mathcal{X}}(\Gamma_n)$ is continuous. Therefore, $\boldsymbol{T}$ is measurable map. In addition, since $y = \boldsymbol{T}(x)$ π-a.e. in $\mathcal{X} \times \mathcal{Y}$, we have for a test function $\varphi \in C_c^\infty(\mathcal{X} \times \mathcal{Y})$ that

$$\int_{\mathcal{X}\times\mathcal{Y}} \varphi(x,y)d\pi(x,y) = \int_{\mathcal{X}\times\mathcal{Y}} \varphi(x,\boldsymbol{T}(x))d\pi(x,y) = \int_{\mathcal{X}} \varphi(x,\boldsymbol{T}(x))d\mu(x).$$

Thus, we have that $\pi = (\mathsf{Id} \times \boldsymbol{T})_{\#}\mu$. □

The above proposition witnesses the interplay of optimal transport maps and plans. We have seen that optimal transport maps are concentrated on c-cyclically monotone sets, which can be obtain by the c-supperdifferential of a c-concave function (or c-subdifferential of a c-convex function respectively). This fact is reflected to the question about whether the c-superdifferential of a c-concave (or again c-subdifferential of a c-convex function resp.) consists a singleton.

Unfortunately, this question has not any answer in general setting. Although, there are many interesting cases of specific problems in different settings. These cases were examinated through the last decades by many authors (see e.g. Smith & Knott (1987); Brenier (1991); Gangbo & McCann (1996); L. Ambrosio et al. (2004); Champion et al. (2008)).

One of the most important of such a case, is the situation of Euclidean spaces and quadratic cost functions. This case, which was mainly explored by C.S. Smith, M. Knott and Y. Brenier and others, providing us various interesting characterisations of optimal transport maps, in terms of Convex Analysis. Such a characterisation is presented in the following proposition.

Proposition 3.4.2. *Consider that $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $c(x,y) = \frac{1}{2}|x-y|^2$ and $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$. Then the following hold:*

(i) we have that φ is c-concave if and only if the map

$$x \mapsto \overline{\varphi}(x) := \frac{|x|^2}{2} - \varphi(x)$$

is convex and lower semicontinuous,

(ii) we have that $y \in \partial^{c+}\varphi(x)$ if and only if $y \in \partial\phi^{c-}\overline{\varphi}(x)$.

Proof. (i) Let us assume that ϕ is c-concave. We will construct $\overline{\phi}$ by equivalent operations and then we will conclude that it is convex and lower semicontinous, by its form.

By definition of c-concavity, we have that $\phi = \psi^{c+}$ for a function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$. This this is equivalent to

$$\begin{aligned}\phi(x) = \inf_{y\in\mathcal{Y}} \{c(x,y) - \psi(y)\} &= \inf_{y\in\mathcal{Y}} \left\{\frac{1}{2}|x-y|^2 - \psi(y)\right\} \\ &= \inf_{y\in\mathcal{Y}} \left\{\frac{|x|^2}{2} + \langle x, -y\rangle + \left(\frac{|y|^2}{2} - \phi(y)\right)\right\},\end{aligned}$$

which is equivalent to

$$\phi(x) - \frac{|x|^2}{2} = \inf_{y\in\mathcal{Y}} \left\{\langle x, -y\rangle + \left(\frac{|y|^2}{2}\psi(y)\right)\right\}$$

and thus, changing the signs and using again c-concavity, is equivalent to

$$\overline{\phi}(x) = \sup_{y\in\mathcal{Y}} \left\{\langle x, y\rangle - \left(\frac{|y|^2}{2} - \phi(y)\right)\right\},$$

which is convex and lower semicontinous, as the Fenchel-Legendre transform of a convex function.
(ii) Let $y \in \partial^{c+}\phi(x)$. Then, by definition we have that this is equivalent to

$$\begin{aligned} \phi(x) &= c(x,y) - \phi^{c+}(y) \\ \phi(z) &\leq c(z,y) - \phi^{c+}(y), \text{ for every } z \in \mathbb{R}^d. \end{aligned}$$

These two relations can be translated in terms of cost function c as

$$\left\{ \begin{aligned} \phi(x) = \frac{|x-y|^2}{2} - \phi^{c+}(y) \\ \phi(z) \leq \frac{|z-y|^2}{2} - \phi^{c+}(y) \end{aligned} \right\} \Longleftrightarrow \left\{ \begin{aligned} \phi(x) - \frac{|x|^2}{2} &= \langle x, -y\rangle + \frac{|y|^2}{2} - \phi^{c+}(y) \\ \phi(z) - \frac{|z|^2}{2} &= \langle z, -y\rangle + \frac{|y|^2}{2} - \phi^{c+}(y) \end{aligned} \right\},$$

for every $z \in \mathbb{R}^d$. Combining them, we conclude that

$$\phi(z) - \frac{|z|^2}{2} \leq \phi(x) - \frac{|x|^2}{2} + \langle z - x, -y\rangle, \quad \text{for every} z \in \mathbb{R}^d.$$

This means that $-y \in \partial^{c+}(\phi - |\cdot|^2)(x)$, or equivalently, $y \in \partial^{c-}\overline{\phi}(x)$. This fact makes our proof complete. □

Before we proceed to discussion of the above result, let us present a standard (and extremely useful) consequence of it.

Corollary 3.4.3 (Special pertubations preserves optimality). *If $\varphi \in C_c^\infty(\mathbb{R}^d)$, then there exists $\overline{\epsilon} > 0$ such that the map $(\mathsf{Id} + \epsilon\nabla\varphi)$ is optimal transport map, for any $|\epsilon| \leq \overline{\epsilon}$.*

Proof. We pick up $\overline{\epsilon} > 0$ such that $-\mathsf{Id} \leq \overline{\epsilon}\nabla^2\varphi \leq \mathsf{Id}$. Under this consideration, the map

$$x \mapsto \frac{|x|^2}{2} + \epsilon\varphi(x), \quad \text{for } |\epsilon| \leq \overline{\epsilon}.$$

Therfore, its gradient is optimal transport map □

The above proposition and its consequence shows up a very important phenomenon which occurs in the Euclidean case under the consideration of quadratic cost. To be more precise, in such a case the condition of concentration on the c-superdifferential of a c-concave map means is translated to the fact that we optimal transport map has to be concentrated on the graph of the subdifferential of a convex function. Moreover, the above proposition reduce the problem of existence of optimal transport maps to the structure of the set of non-differentiability points of a convex function. Differentiability of convex functions has been widely studied last half century under different prespectives and it is appearing many extraordinary results. In order to understand this set in our case, we recall the definition of (CC)- hypersurfaces[3]

Definition 3.4.4 ((CC)-hypersurfaces). Let $\mathscr{O} \subset \mathbb{R}^d$. If for a suitable system of coordinates, this set is the graph of the difference of two real-valued fuctions, we will say that $\mathscr{O}$ is a (CC)-hypersurface. In the other words, the set $\mathscr{O}$ is a (CC)-hypersurfaces, if there exist to convex functions $f, g : \mathbb{R}^{d-1} \to \mathbb{R}$, such that

$$\mathscr{O} := \left\{(s,t) \in \mathbb{R}^d : s \in \mathbb{R}^{d-1}, t \in \mathbb{R}, t = f(s) - g(s)\right\}.$$

[3] The notation (CC) stands for *convex minus convex,* following Ambrosio L., Gigli N. (2013), and it has nothing to do about the cost function c. Such a hypersurfaces where introduced in Zajíček (1979) and appearing in many situations (see e.g. Borwein et al. (2010); Benyamini & Lindenstrauss (1998).

Having in our minds the notion of (CC)-hypersurfaces, the break points of non-differentiability of a convex function and moreover the structure of their set are described in the following proposition.

Proposition 3.4.5 (Zajíček (1979)). *If $\mathscr{O} \subset \mathbb{R}^d$ then there exists a convex function $\overline{\phi} : \mathbb{R}^d \to \mathbb{R}$ such that $\mathscr{O}$ is contained in the set of non-differentiability points of ϕ if and only if $\mathscr{O}$ can be covered by countably many (CC)-hypersurfaces.*

Proof. This proof, due to its technicalities, is omitted Although, it can be found in original paper Zajíček (1979) in general Banach space setting. Moreover, a modern approach can be found in Benyamini & Lindenstrauss (1998). □

Corollary 3.4.6 (Regularity of measures on (CC)-hypersurfaces). *If a measure $\mu \in \mathscr{P}(\mathbb{R}^d)$ is regular, then for every (CC)-hypersurface $\mathscr{O} \subset \mathbb{R}^d$ holds that $\mu(\mathscr{O}) = 0$.*

Having explore the set of non-differentiability of a convex function, we are ready to state and prove one of the most important result in Euclidean setting under quadratic costs, the famous *Brenier's Theorem.*

Theorem 3.4.7 (Brenier). *Consider a regular measure $\mu \in \mathscr{P}(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d} |x|^2 d\mu(x) < \infty$. Then for every $\nu \in \mathscr{P}(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} |y|^2 d\nu(y) < \infty$, there is a unique transport plan between μ and ν and its induced by a measurable map $\boldsymbol{T} : \mathbb{R}^d \to \mathbb{R}$. Moreover, the optimal transport map $\boldsymbol{T}$ is the gradient of a convex function.*

Proof. We recall the cost condition in Kantorovich problem, which was describted in n describted in Theorem 3.2.16, that is

$$c(x, y) \leq a(x) + b(y) \tag{3.31}$$

We plug in 3.31 $a(x) = b(x) = |x|^2$. Then, thanks to our assumptions on measure $\mu \in \mathscr{P}(\mathbb{R}^d)$, the measure $\nu \in \mathbb{R}^d$ ensure us that the cost condition (3.31). Therefore, both of Theorem 3.2.16 and Theorem 3.3.2 are in our favour. Thanks to relation between duality and c-concavity, which was describted in Remark 3.3.3, we know that for any c-concave potential ϕ and any optimal transport plan $\pi \in \Pi(\mu, \nu)$, it holds that $\operatorname{supp}(\pi) \subset \partial^{c+}\phi$.

Now, Proposition 3.4.2 ensure us that the functional

$$\overline{\phi}(\cdot) := \frac{|\cdot|^2}{2} - \phi(\cdot)$$

is convex, and moreover, it holds that $\partial^c \phi = \partial^{c-}\overline{\phi}$. Using again our assumptions on $\mu \in \mathscr{P}(\mathbb{R}^d)$ and since $\overline{\phi}$ is convex, we know that the set $\mathscr{O}$ of the points of nondifferentiability of $\overline{\phi}$ is μ-null set. Hence, the map $\nabla\overline{\phi} : \mathbb{R}^d \to \mathbb{R}^d$ is well-defined μ-a.e. and every transport plan has to be concentrated on its graph. Thus, the optimal transport plan is unique and it is induced by the gradient of the function $\overline{\phi}$, which as we have seen is convex. This fact makes our proof complete. □

CHAPTER 4

Wasserstein spaces

"I prefer concrete things and I don't like to learn more about abstract stuff than I absolutely have to.
"

— Mark Kac,
Polish-American mathematician (1914-1984)

In this chapter, we will focus on the optimal transport problem under a specific cost function, which will be defined through a distance in a metric space, i.e. $c(x, y) = d^p(x, y)$, for $p \in [1, \infty)$. This consideration gives rise to the notion of so-called *Wasserstein distance* and *Wasserstein spaces* which naturally appears from the optimal transport problem and it presents a plethora of interesting and important topological and geometrical properties. Moreover, based on those properties we will recast the dynamical formulation of the optimal transport problem, which will play a crucial role to understanding many geometrical features of Wasserstein spaces.

4.1 Preliminaries and background

It is well-known that the set of Borel probability measures on a complete and separable metric space can be endowed with the topology of narrow convergence (see e.g. Parthasarathy (2005); Aliprantis & Border (2006)). This topology is a metrizable topology for several metrics, as for example the bounded Lipschitz metric, the Lévy-Prokhorov metric or many others[1]. Heuristically speaking, having the transportation problem between two probability measures, which was described in the previous chapter, one can somehow understand its minimal value as a vague notion of distance between probability measures. This heuristic and vague consideration might not be true in general for arbitrary cost functions, nevertheless, if one considers a distance as a cost function, optimal transportation problem, as we will see in the following, can provide us with a true notion of distance with respect to the topology of narrow convergence in the space of probability measures, the so-called *Wasserstein distances.*

During the 20th century, Wasserstein-type distances were discovered and rediscovered under different names, by many authors in different settings, and by different scientific communities. Even the so-called term *Wasserstein* maybe is a debatable, enjoying a colorful history. More precisely,

[1]The number of such metrics is surprisingly large! An excellent treatise about these definitions and the metric methods in Probability and Statistics can be found in Rachev et al. (2013).

it was R.L Dobrushin who introduced that term (see Dobrushin (1970)), based on the work of Russian, US-based, mathematician L. Vaserstein (see Vaserstein (1969)). Essentially, the notion of such distances does not appearing explicitly in the work of Vasserstein (or possibly respelled as Wasserstein), and probably he actually didn't play any key role that could on the development of the theory of these distances. Despite that fact, there are many names in the modern literature referring to the same concept, such as minimal L^p-distance (established by Rüschendorf and Rachev), Earth Mover's distance (established in Computer Science literature), Mallow's distance (established in Probability and Statistics literature), Kantorovich-Rubinstein distance and possibly many others. Nowadays, the term Wasserstein is an undistinguished and widely accepted terminology in a wide range of scientific disciplines[2], from differential geometry and partial differential equations to optimization and machine learning, and therefore it may be impossible to change.

The aim of this chapter is to provide a gentle investigation of the topological and geometrical properties of Wasserstein distances, and moreover, of the Wasserstein spaces. As we will see, the notion of Wasserstein distances enjoy several interesting properties, which essentially lies to the topological and the geometrical structure of its base space.

4.2 Metric nature and topology

We are starting our trip to Wasserstein spaces by an investigation of the metric nature and its topological properties, which we have already advertised. We shall start this investigation by considering that $(\mathcal{X}, d)$ is a complete and separable metric space. We define the p-Wasserstein distance between $\mu, \nu \in \mathscr{P}(\mathcal{X})$ as

$$\mathcal{W}_p(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d^p(x,y)d\pi(x,y) \right)^{1/p}. \tag{4.1}$$

As we have already mentioned, the quantity defined in (4.1) is a true distance in the space of probability measures $\mathscr{P}(\mathcal{X})$. In order to prove that claim, we recall the famous gluing lemma, which makes a clever use of the disintegration theorem (see L. Ambrosio et al. (2008)).

Lemma 4.2.1 (Gluing). *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be complete and separable metric spaces and two probability measures $\pi_1 \in \mathscr{P}(\mathcal{X}\times\mathcal{Y})$ and $\pi_2 \in \mathscr{P}(\mathcal{Y}\times\mathcal{Z})$ such that $(\mathsf{p}_{\mathcal{Y}})_{\#}\pi_1 = (\mathsf{p}_{\mathcal{Y}})_{\#}\pi_2$. Then there exists a measure $\pi \in \mathscr{P}(\mathcal{X}\times\mathcal{Y}\times\mathcal{Z})$ such that*

$$(\mathsf{p}_{\mathcal{X},\mathcal{Y}})_{\#}\pi = \pi_1 \text{ and } (\mathsf{p}_{\mathcal{Y},\mathcal{Z}})_{\#}\pi = \pi_2$$

Proof. Consider $\mu = (\mathsf{p}_{\mathcal{Y}})_{\#}\pi_1 = (\mathsf{p}_{\mathcal{Y}})_{\#}\pi_2$. Using the disintegration theorem, we can write that

$$d\pi_1(x,y) = d\mu(y)d\pi_1^y \text{ and } d\pi_2(y,z) = d\mu(y)d\pi_2^y(z).$$

We conclude the proof defining π as

$$d\pi(x,y,z) := d\mu(y)d(\pi_1^y \times \pi_2^y)(x,z).$$

□

Now, having already developed the necessary machinery, we are ready and able to prove that the quantity (4.1) is a true distance on $\mathscr{P}(\mathcal{X})$ in the following proposition.

[2]Many people quoting this fact as *Wassersteinization,* i.e. introducing optimal transportation methods into an optimization or machine learning problem.

Proposition 4.2.2 ($\mathcal{W}_p$ is a distance on $\mathscr{P}(\mathcal{X})$). *The p-Wasserstein distance $\mathcal{W}_p$ satisfies the axioms of a distance on $\mathscr{P}(\mathcal{X})$, i.e. if $\mu_1, \mu_2, \mu_3 \in \mathscr{P}(\mathcal{X})$ then*

(i) $\mathcal{W}_p(\mu_1, \mu_2) \geq 0$ *(non negativity)*

(ii) $\mathcal{W}_p(\mu_1, \mu_2) = \mathcal{W}_p(\mu_2, \mu_1)$ *(symmetry)*

(iii) $\mathcal{W}_p(\mu_1, \mu_2) = 0$ *implies that* $\mu_1 = \mu_2$ *(identity of indiscernibles)*

(iv) $\mathcal{W}_p(\mu_1, \mu_3) \leq \mathcal{W}_p(\mu_1, \mu_2) + \mathcal{W}_p(\mu_2, \mu_3)$. *(triangle inequality)*

Proof. The claims $(i), (ii)$ are straightforward.
For (iii), we pick an optimal transport plan π and we observe that $\int_{\mathcal{X}\times\mathcal{X}} d^p(x,y)d\pi(x,y) = 0$ implies that π is concentrated on the diagonal of $\mathcal{X} \times \mathcal{X}$, which means that the two canonical projection maps $\mathsf{p}_1, \mathsf{p}_2$ coincide π-a.e. Thus, we have that $(\mathsf{p}_1)_{\#}\pi = (\mathsf{p}_2)_{\#}\pi$.

For (iv), we will use the Gluing lemma, in order to construct two optimal transport plans. Let $\mu_1, \mu_2, \mu_3 \in \mathscr{P}(\mathcal{X})$ and let $\pi_{1,2}$ be the optimal transport plan between μ_1 and μ_2 and $\pi_{2,3}$ be the optimal transport plan between μ_2, μ_3. Thanks to the Gluing lemma we know that there exists $\pi \in \mathscr{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{X})$ such that

$$(\mathsf{p}_{1,2})_{\#}\pi = \pi_{1,2} \text{ and } (\mathsf{p}_{2,3})_{\#}\pi = \pi_{2,3}.$$

Since $(\mathsf{p}_1)_{\#}\pi = \mu_1$ and $(\mathsf{p}_3)_{\#}\pi = \mu_3$, we have that $(\mathsf{p}_{1,3})_{\#}\pi \in \Pi(\mu_1, \mu_3)$ and therefore from the triangle inequality in $L^p(\pi)$, we obtain

$$\begin{aligned}
\mathcal{W}_p(\mu_1, \mu_3) &\leq \left(\int_{\mathcal{X}\times\mathcal{X}} d^p(x_1, x_2) d(\mathsf{p}_{1,3})_{\#}\pi(x_1, x_3)\right)^{1/p} = \left(\int_{\mathcal{X}\times\mathcal{X}\times\mathcal{X}} d^p(x_1, x_3) d\pi(x_1, x_2, x_3)\right)^{1/p} \\
&\leq \left(\int_{\mathcal{X}\times\mathcal{X}\times\mathcal{X}} \big(d(x_1, x_2) + d(x_2, x_3)\big)^p d\pi(x_1, x_2, x_3)\right. \\
&\leq \left(\int_{\mathcal{X}\times\mathcal{X}\times\mathcal{X}} d^p)(x_1, x_2) d\pi(x_1, x_2, x_3)\right)^{1/p} + \left(\int_{\mathcal{X}\times\mathcal{X}\times\mathcal{X}} d^p(x_2, x_3) d\pi(x_1, x_2, x_3)\right)^{1/p} \\
&= \left(\int_{\mathcal{X}\times\mathcal{X}\times\mathcal{X}} d^p(x_1, x_2) d\pi_{1,2}(x_1, x_2)\right)^{1/p} + \left(\int_{\mathcal{X}\times\mathcal{X}\times\mathcal{X}} d^p(x_2, x_3) d\pi_{2,3}(x_2, x_3))\right)^{1/p} \\
&= \mathcal{W}_p(\mu_1, \mu_2) + \mathcal{W}_p(\mu_2, \mu_3).
\end{aligned}$$

□

To make sure that $\mathcal{W}_p$ is finite, we restrict ourselves to the space of propability measures with p-finite moments, i.e. $\mathscr{P}(\mathcal{X})$. In this spirit, we define the Wasserstein space as follows.

Definition 4.2.3 (Wasserstein space). Consider that $(\mathcal{X}, d)$ is a complete and separable metric space. We define the $p-$Wasserstein space as the space of probability measures with finite p-moments equipped with p-Wasserstein distance, that is

$$\mathbb{W}_p(\mathcal{X}) := (\mathscr{P}_p(\mathcal{X}), \mathcal{W}_p) = \left\{\mu \in \mathscr{P}(\mathcal{X}) : \int_{\mathcal{X}} d^p(x_0, x) d\mu < \infty, \text{ for some } x_0 \in \mathcal{X}\right\}.$$

Therefore, by now, the space of probability measures with p-finite moments enjoys all of the classical metric properties that we know from Analysis. In what will follow, we will focus in the case where $p = 2$.

The reason of doing this threefold: firstly, the case where $p = 2$ enjoys a interesting and nice geometrical flavour, secondly, it is also appearing in many applications, in an natural way, for example, it reassembles notion the covariance of two random variables in probability theory, and thirdly most of the results which will follow can be generalized for general $p \in [1, +\infty)$ by minor changes.

As we mention in the introductory paragraph of this chapter, there are many ways to metrize the space of probability measures. Probably, the simplest and more classical way to do this is using the total variation between two probability measures $\mu, \nu \in \mathscr{P}_p(\mathcal{X})$, i.e. the quantity

$$\|\mu - \nu\|_{TV} := |\mu - \nu|(\mathcal{X}) = 2 \sup \big\{ |\mu(A) - \nu(A)| : A \in \mathcal{B}(\mathcal{X}) \big\}.$$

Although,the later notion of distance in space of probability measures may be sightly dangerous. To be more precise, the total variation distance in several situations *cannot* measure very large or very small distances, and admits several pathologies, such as the case where μ, ν are Dirac measures. Nevertheless, having the metric nature of Wasserstein space $\mathbb{W}_2(\mathcal{X})$ in our mind, an interesting question to ask is: what about the relation of total variation and Wasserstein distance in the space of probability measures? A first understanding of such a concern comes from the probabilistic representation of total variation. In particular, having two random variables X, Y where $X \sim \mu$, and $Y \sim \nu$, we know that total variation between μ and ν, can be translated in the following nice formula

$$\|\mu - \nu\|_{TV} = 2 \inf \mathbb{P}(X \neq Y),$$

where the infimum is taken along joint probability distributions (X, Y) having X, Y as marginals. This formulation reads as a very particular case of Kantorovich duality, where was discussed in previous section, by considering as cost function the indicator $\mathbb{1}_{X \neq Y}$.

Based on this observation, controlling some quantity in Wasserstein distance seems to be much more weaker than controlling the same quantity in total variation. Nevertheless, this intuitive consideration it is not completely true, since as we mention above total variation does not cares about possible huge or tiny distances between measures. In any case, the good news, is that Wasserstein distance can be controlled by a weighted version of total variation, as is presented in the following proposition.

Proposition 4.2.4 (Total variation bound)**.** *Let $(\mathcal{X}, d)$ be a complete and separable metric space, two probability measures $\mu, \nu \in \mathscr{P}_2(\mathcal{X})$ and $z \in \mathcal{X}$. Then it holds that*

$$\mathcal{W}_2(\mu, \nu) \leq 2^{1/2} \left(\int_{\mathcal{X} \times \mathcal{X}} d^2(z, x) d|\mu - \nu|(x) \right)^{1/2} \tag{4.2}$$

Proof. Let $\pi \in \Pi(\mu, \nu)$ be a transport plan, which is obtained by fixing the mass shared by μ and ν and distributing the rest uniformly, that is

$$\pi := (\mathsf{Id}, \mathsf{Id})_{\#}(\mu \wedge \nu) + \frac{1}{\alpha}(\mu - \nu)_+ \otimes (\mu - \nu)_-,$$

where $\mu \wedge \nu = \mu - (\mu - \nu)_+$ and $\alpha = (\mu - \nu)_-(\mathcal{X}) = (\mu - \nu)_+(\mathcal{X})$. Then, by definition of Wasserstein distance, the definition of transport plan π, the triangle inequality for metric d, the above relation for α and using the simple inequality $(A+B)^2 \leq 2^{1/2}(A^2+B^2)$ for $A, B \geq 0$, we have

$$\begin{aligned}
\mathcal{W}_2^2(\mu,\nu) &\leq \int_{\mathcal{X}\times\mathcal{X}} d^2(x,y)d\pi(x,y) \\
&= \frac{1}{\alpha}\int_{\mathcal{X}}\int_{\mathcal{X}} d^2(x,y)d(\mu-\nu)^+(x)d(\mu-\nu)^-(y) \\
&\leq \frac{2}{\alpha}\int_{\mathcal{X}\times\mathcal{X}} \big(d^2(x,z)+d^2(z,y)d(\mu-\nu)^+(x)d(\mu-\nu)^-(y) \\
&\leq 2\Big(\int_{\mathcal{X}} d^2(x,z)d(\mu-\nu)^+ + \int_{\mathcal{X}} d^2(z,y)d(\mu-\nu)^-\Big) \\
&= 2\int_{\mathcal{X}\times\mathcal{X}} d^2(x,z)d\big((\mu-\nu)^+ + (\mu-\nu)^-\big)(x) \\
&= 2\int_{\mathcal{X}\times\mathcal{X}} d^2(x,z)d|\mu-\nu|(x).
\end{aligned}$$

Thus, we have

$$\mathcal{W}_2(\mu,\nu) \leq 2^{1/2}\Big(\int_{\mathcal{X}\times\mathcal{X}} d^2(z,x)d|\mu-\nu|(x)\Big)^{1/2},$$

which makes our proof complete. □

One might observe that above weighted bound of total variation coincides with the distance $\mathcal{W}_1$, when the cost function is given by $c(x,y) := (d(x,z)+d(z,y)\mathbb{1}_{x\neq y}$. This observation will play a crucial role in order to investigate topological properties of $\mathbb{W}_2(\mathcal{X})$ in a deeper level, as we will see in the following.

Having established the metric structure of $\mathbb{W}_2(\mathcal{X})$ and explored a specific weighted bound for Wasserstein distances, we will focus our attention to a class of certain and finer topological properties. In order to investigate such properties, we recall the notions of *quadratic growth* and *2-uniform integrability.*

Property 4.2.5 (Quadratic growth)**.** We remind that a functional $f : \mathcal{X} \to \mathbb{R}$ has quadratic growth if

$$|f(x)| \leq a(d^2(x,x_0)+1) \tag{4.3}$$

for some $a \in \mathbb{R}$ and some $x_0 \in \mathcal{X}$.

The property of quadratic growth of a certain functional with play a crucial role in the following, translated as the convergence of 2-moment of probability measures. An important consequence of this definition is that if f has quadratic growth and $\mu \in \mathscr{P}_2(\mathcal{X})$, then $f \in L^1(\mathcal{X},\mu)$.

Property 4.2.6 (2-uniform integrability)**.** For given $\epsilon > 0$ and some $x_0 \in \mathcal{X}$, we will say that $K \subset \mathscr{P}_2(\mathcal{X})$ is 2-uniformly integrable if there exists $R_\epsilon > 0$ such that

$$\sup_{\mu\in K}\int_{\mathcal{X}\setminus B_{R_\epsilon}(x_0)} d^2(x,x_0)d\mu \leq \epsilon.$$

The property of 2-uniform integrability has a nice important collateral consequence, which will be also very useful in the following. To be more accurate, if $\mathcal{X},\mathcal{Y}$ are complete and separable metric spaces, one can prove that if $K_1 \subset \mathscr{P}_2(\mathcal{X})$ and $K_2 \subset \mathscr{P}_2(\mathcal{Y})$ are 2-uniformly integrable, then the set

$$\big\{\pi \in \mathscr{P}(\mathcal{X}\times\mathcal{Y}) : (\mathsf{p}_{\mathcal{X}})_{\#}\pi \in K_1 \text{ and } (\mathsf{p}_{\mathcal{Y}})_{\#}\pi \in K_2\big\}$$

is also 2-uniformly integrable.

Note that these notions become also meaningful for general $p \in [1, +\infty)$.

Moreover, intuitively, 2-uniform integrability is a compactness-type concept, which will be vital in the following. Precisely, combined with the convergence of integrals of functionals with quadratic growth, it will make us able to resembling the tightness combined with the convergence of integrals of bounded functions.

The following proposition witness this connection, and moreover, describes and provides a first characterization of the convergence notion in Wasserstein space $\mathbb{W}_2(\mathcal{X})$.

Proposition 4.2.7. *Let $(\mu_n)_{n\in\mathbb{N}} \in \mathscr{P}_2(\mathcal{X})$ be a sequence of probability measures narrowly converging to some μ. Then the following statements are equivalent:*

(i) the sequence $(\mu_n)_{n\in\mathbb{N}}$ is 2-uniformly integrable

(ii) $\displaystyle\int_{\mathcal{X}} f d\mu_n \longrightarrow \int_{\mathcal{X}} f d\mu$, as $n \to \infty$ for any continuous function f with quadratic growth

(iii) $\displaystyle\int_{\mathcal{X}} d^2(\cdot, x_0) d\mu_n \longrightarrow \int_{\mathcal{X}} d^2(\cdot, x_0) d\mu$, as $n \to \infty$ for some $x_0 \in \mathcal{X}$.

Proof. $(i) \Rightarrow (ii)$: Without loss of generality, we assume that $f \geq 0$. Based on this fact, f can be written as supremum of a family of continuous and bounded functions, and clearly we have that

$$\int_{\mathcal{X}} f d\mu \leq \liminf_{n\to\infty} \int_{\mathcal{X}} f d\mu_n.$$

Therefore, it suffices to prove is the limsup inequality, in order to obtain the limit. For that reason, we fix $\epsilon > 0$, $x_0 \in \mathcal{X}$ and we find a radius $R_\epsilon > 1$ such that

$$\int_{\mathcal{X}\setminus B_{R_\epsilon}(x_0)} d^2(\cdot, x_0) d\mu_n \leq \epsilon, \text{ for every } n.$$

Now, if χ is the function with bounded support, values in $[0, 1]$ and identically equal with 1 in B_{R_ϵ}, then for every $n \in \mathbb{N}$ and $a \in \mathbb{R}$, we obtain

$$\int_{\mathcal{X}} f d\mu_n = \int_{\mathcal{X}} f\chi d\mu_n + \int_{\mathcal{X}} f(1-\chi) d\mu_n \leq \int_X f\chi d\mu_n + \int_{X\setminus B_{R_\epsilon}} f d\mu_n \leq \int_{\mathcal{X}} f\chi d\mu_n + 2a\epsilon.$$

Since the function $f\chi$ is continuous and bounded, we have that $\int_{\mathcal{X}} f\chi d\mu_n \to \int_{\mathcal{X}} f\chi d\mu$, thus we obtain

$$\limsup_{n\to\infty} \int_{\mathcal{X}} f d\mu_n \leq \int_{\mathcal{X}} f\chi d\mu + 2a\epsilon \leq \int_{\mathcal{X}} f d\mu + 2a\epsilon.$$

Thanks to the arbitrariness of $\epsilon > 0$, we get also the limsup inequality, so our thesis is proved.
$(ii) \Rightarrow (iii)$: Since d^2 are trivially continuous with quadratic growth the proof is straightforward.
$(iii) \Rightarrow (ii)$: Suppose for the sake of contradiction that there $\epsilon > 0$ and $\tilde{x}_0 \in \mathcal{X}$ such that for every $R > 0$, it holds that

$$\sup_{n\in\mathbb{N}} \int_{\mathcal{X}\setminus B_R(\tilde{x})} d^2(\cdot, \tilde{x}_0) d\mu_n > \epsilon.$$

Then, it is not hard to see that it also holds

$$\limsup_{n\to\infty} \int_{\mathcal{X}\setminus B_R(\tilde{x}_0)} d^2(\cdot, x_0) d\mu_n > \epsilon. \tag{4.4}$$

Now, for every $R > 0$, let χ_R be a continuous cut-off function with values in $[0,1]$, supported on $B_R(\tilde{x}_0)$ and be identically equal with 1 on $B_{R/2}(\tilde{x}_0)$. Since $d^2(\cdot,\tilde{x}_0)\chi_R$ is continuous and bounded, we obtain

$$\begin{aligned}
\int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)\chi_R d\mu &= \lim_{n\to\infty}\int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)\chi_R d\mu_n \\
&= \lim_{n\to\infty}\left(\int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)d\mu_n - \int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)(1-\chi_R)d\mu_n\right) \\
&= \int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)d\mu - \lim_{n\to\infty}\int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)(1-\chi_R)d\mu_n \\
&\leq \int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)d\mu - \liminf_{n\to\infty}\int_{\mathcal{X}\setminus B_R(\tilde{x}_0)} d^2(\cdot,\tilde{x}_0)d\mu_n \\
&= \int_{X} d^2(\cdot,\tilde{x}_0)d\mu - \limsup_{n\to\infty}\int_{\mathcal{X}\setminus B_R(\tilde{x}_0)} d^2(\cdot,\tilde{x}_0)d\mu_n \\
&\overset{(4.4)}{\leq} \int_{\mathcal{X}} d^2(\cdot,x_0)d\mu - \epsilon.
\end{aligned}$$

On the other hand, we have

$$\int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)d\mu = \sup_{R>0}\int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)\chi_R d\mu \leq \int_{\mathcal{X}} d^2(\cdot,\tilde{x}_0)d\mu - \epsilon,$$

a contradiction, which concludes our proof. □

Having by now a first characterization of the meaning of convergence in Wasserstein space $\mathbb{W}_2(\mathcal{X})$, we proceed our investigation by exploring a important property of Wasserstein distance $\mathcal{W}_2$, that is *lower semicontinuity.* As a byproduct of that, we can obtain that a sequence of optimal transport plans are *stable,* that is, the limit of a sequence of optimal transport plans is still optimal. Both of these claims are presented in the following proposition.

Proposition 4.2.8 (Lower semicontinuity of $\mathcal{W}_2$ and stability of optimality)**.** *Consider that $(\mathcal{X},d)$ is a complete and separable metric space. Then the distance $\mathcal{W}_2$ is lower semicontinuous with respect to the topology of narrow convergence of measures. Moreover, if $(\pi_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathcal{X}\times\mathcal{X})$ is a sequence of optimal transport plans, its narrow limit $\pi \in \mathscr{P}_2(\mathcal{X}\times\mathcal{X})$ is also optimal transport plan.*

Proof. At first, let us prove lower semicontinuity of $\mathcal{W}_2$. We consider two sequences of probability measures $(\mu_n)_{n\in\mathbb{N}}, (\nu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathcal{X})$ having their narrow limits $\mu,\nu \in \mathscr{P}_2(\mathcal{X})$ respectively. We pick a sequence of optimal transport plans $(\pi_n)_{n\in\mathbb{N}}$. Thanks to Proposition 3.2.5, we know that the set of transport plans is tight, and moreover, thanks to Proposition 3.2.4 (Prokhorov's theorem), the sequence $(\pi_n)_{n\in\mathbb{N}}$ admits a subsequence, which is converging narrowly to some $\pi \in \mathscr{P}_2(\mathcal{X}\times\mathcal{X})$. Thusly, it is clear that $(\mathsf{p}_1)_{\#}\pi = \mu$ and $(\mathsf{p}_2)_{\#}\pi = \nu$, and hence it holds that

$$\mathcal{W}_2^2(\mu,\nu) \leq \int_{\mathcal{X}\times\mathcal{X}} d^2(x,y)d\pi(x,y) \leq \liminf_{n\to\infty}\int_{\mathcal{X}\times\mathcal{X}} d^2(x,y)d\pi_n(x,y) = \liminf_{n\to\infty}\mathcal{W}_2^2(\mu_n,\nu_n).$$

So, we have

$$\mathcal{W}_2^2(\mu,\nu) \leq \liminf_{n\to\infty}\mathcal{W}_2(\mu_n,\nu_n),$$

which proves the lower semicontinuity of $\mathcal{W}_2$.

Now, let us prove the stability of optimality. Using the same notation, the only thing we have to prove that the first statement holds true and that the narrow limit $\pi \in \mathscr{P}(\mathcal{X}\times\mathcal{X})$ is also optimal. To do this, we choose $a(x) = b(x) = d^2(x, x_0)$ for some $x_0 \in \mathcal{X}$ satisfying the cost condition $c(x) \leq a(x) + b(x)$ as it given in Theorem 3.2.16. Under this perspective, and since $\mu, \nu \in \mathscr{P}_2(\mathcal{X})$, it is not hard to observe that Theorem 3.2.16 can be applied, thus the optimality of π is equivalent to c-monotonicity of $\text{supp}(\pi)$. The same fact also holds for the transport plans π_n. Under this considerations, we fix a $N \in \mathbb{N}$ and we pick a sequence $(x^i, y^i)_{i=1}^N \in \text{supp}(\pi)$. Since $(\pi_n)_{n\in\mathbb{N}}$ converges to π in narrow sense, one can infer the existence of $(x_n^i, y_n^i)_{n\in\mathbb{N}} \in \text{supp}(\pi_n)$ such that

$$\lim_{n\to\infty} \big(d(x_n^i, x^i) + d(y_n^i, y^i)\big) = 0, \text{ for } i = 1, \cdots, N.$$

So, thanks to the c-monotonicity of $\text{supp}(\pi_n)$ and since in our case the cost function is continuous, our proof is completed. □

We proceed now to a further topological investigation related to sequence in Wasserstein spaces.

Proposition 4.2.9 (Cauchy means tight)**.** *Consider that $(\mathcal{X}, d)$ is a complete and separable metric space and $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_p(\mathcal{X})$ is a Cauchy sequence with respect to $\mathcal{W}_2$. Then $(\mu_n)_{n\in\mathbb{N}}$ is tight.*

Proof. In order to proof this proposition, we will be based in the following purely topological strategy. At first, we will employ Cauchy property of the corresponding sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ and the duality formula, finding a family of sets where this sequence gives "small" mass. Then, since there is to guarantee that this family is compact, we will find a "nice" compact replacement of that family, and the desired result will follow. For a better exposition of this strategy, we divide the proof into 2 steps.

Step 1: *Finding the family of "small" sets.*
Consider that $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathcal{X})$ is a Cauchy sequence. This means that $\mathcal{W}_2(\mu_n, \mu_m) \to 0$, as $n, m \to \infty$, and moreover, playing with the terms for some $x_0 \in \mathcal{X}$, we have

$$\int_{\mathcal{X}} d^2(x_0, x) d\mu_n(x) = \mathcal{W}_2^2(\delta_{x_0}, \mu_n) \leq \Big(\mathcal{W}_2(\delta_{x_0}, \mu_1) + \mathcal{W}_2(\mu_1, \mu_n)\Big)^2 < \infty,$$

as $n \to \infty$. Now, by a standard application of Hölder's inequality, we can see that $\mathcal{W}_1 \leq \mathcal{W}_2$, and thusly, the sequence $(\mu_n)_{n\in\mathbb{N}}$ is also Cauchy with respect to $\mathcal{W}_1$.

Let $\epsilon > 0$ and $N \in \mathbb{N}$, such that for every $n \geq N$, we have

$$W_1(\mu_n, \mu_N) < \epsilon^2.$$

Then, for any $n \in \mathbb{N}$, there is $i \in \{1, \cdots, N\}$ such that

$$W_1(\mu_i, \mu_N) < \epsilon^2. \tag{4.5}$$

Note, that in the case where $n < N$, the bound (4.5) also holds, by choosing $i = n$. Now, since by construction, the finite set $\{\mu_1, \cdots, \mu_N\}$ is tight, there exists a compact set K such that $\mu_i(\mathcal{X} \backslash K) < \epsilon$ for any $i \in \{1, \cdot N\}$. Therefore, due to its compactness, K can be covered by finitely many balls with centers x_i and radius ϵ, that is

$$K \subset \bigcup_{i=1}^{\ell} B(x_i, \epsilon).$$

We define the sets

$$\mathcal{U} := \bigcup_{i=1}^{\ell} B(x_i, \epsilon), \quad \mathcal{U}_\epsilon := \{x \in \mathcal{X} : d(x, \mathcal{U}) < \epsilon\}$$

and the map

$$\phi(x) := \Big(1 - \frac{d(x,\mathcal{U})}{\epsilon}\Big)^+.$$

From the above definitions, we can observe that the function ϕ enjoys some interesting properties. More precisely, it is bounded from above and below by the indicator functions of $\mathcal{U}$ and $\mathcal{U}_\epsilon$ respectively, i.e. $\mathbb{1}_{\mathcal{U}} \leq \phi \leq \mathbb{1}_{\mathcal{U}_\epsilon}$, and moreover, it is $(1/\epsilon)$-Lipchitz. Recalling the Kantorovich-Rubinstein duality formula of Remark 3.3.4, and employing the above properties of ϕ, for any $i < N$ we have

$$\begin{aligned}\mu_n(\mathcal{U}_\epsilon) &\geq \int_{\mathcal{X}} \phi d\mu_n = \int_{\mathcal{X}} \phi d\mu_i + \Big(\int_{\mathcal{X}} \phi d\mu_n - \int_{\mathcal{X}} d\mu_i\Big)\\ &\geq \int_{\mathcal{X}} \phi d\mu_i - \frac{\mathcal{W}_1(\mu_i,\mu_n)}{\epsilon}\\ &\geq \mu_i(\mathcal{U}) - \frac{\mathcal{W}_1(\mu_i,\mu_n)}{\epsilon}.\end{aligned}$$

Therefore, on the one hand, we have that

$$\mu_i(\mathcal{U}) \geq \mu_i(K) \geq 1 - \epsilon \quad \text{for } i \leq N,$$

and on the other hand, for each $n \in \mathbb{N}$, we can find a $i = i(n)$ such that

$$\mathcal{W}_1(\mu_i,\mu_n) \leq \epsilon^2.$$

Based on these facts, we infer that

$$\mu_n(\mathcal{U}_\epsilon) \geq 1 - \epsilon - \frac{\epsilon^2}{\epsilon} = 1 - 2\epsilon.$$

So, by far, we conclude that for any $\epsilon > 0$, we can find a finitely many $(x_i)_{i=1}^{\ell}$ such that all of the terms of the sequence $(\mu_n)_{n\in\mathbb{N}}$ gives mass less that $1 - 2\epsilon$ to the set

$$C := \bigcup_{i=1}^{\ell} B(x_i, 2\epsilon).$$

In this point, we note that there is no guarantee that the set C is compact. To bypass such a technical difficult, we will use an "approximation" type trick to find a replacement set, which is described in Step 2.

Step 2: *Find a compact replacement.*
To find a compact replacement of C, the only one that we have to do is replace ϵ in Step 1 by $2^{-(s+1)}\epsilon$ with $s \in \mathbb{N}$. Under this consideration, there will be a family $(x_i)_{i=1}^{m(s)}$ such that

$$\mu_n\Big(\mathcal{X}\backslash \bigcup_{i=1}^{m(s)} B(x_i, 2^{-s}\epsilon\Big) \leq 2^{-s}\epsilon.$$

Therefore, we have $\mu_n(\mathcal{X}\backslash S) \leq \epsilon$, where the set S is defined by

$$S := \bigcap_{k=1}^{\infty} \bigcup_{i=1}^{m(k)} \overline{B(x_i, 2^{-k}\epsilon)}.$$

Moreover, by construction, choosing ℓ large enough such that $2^{-\ell} < \delta$, for some $\delta > 0$ arbitrary small, the set S can be covered by finitely many balls radius δ, i.e

$$B(x_i, 2^{-\ell}\epsilon) \subset B(x_i, \delta).$$

In addition, S is closed, as finite intersection of closed sets.

To conclude, since $\mathcal{X}$ is a complete metric space, it follows that S is compact, and $(\mu_n)_{n\in\mathbb{N}}$ is tight, which makes our proof complete. □

Having now the above nice tightness characterization in our machinery, we are ready to prove one of the most important properties of Wasserstein space $\mathbb{W}_2(\mathcal{X})$ in the following proposition.

Theorem 4.2.10 (A fundamental property of $\mathbb{W}_2(\mathcal{X})$)**.** *Coniser that $(\mathcal{X}, d)$ is complete and separable metric space and $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathcal{X})$ be a sequence of probability measures with second finite moments. Then, we have that*

$$\mathcal{W}_2(\mu_n, \mu) \to 0 \text{ if and only if } \mu_n \longrightarrow \mu \text{ in narrow sence.}$$

Proof. We will start by proving the $(\Rightarrow)$ implication.
Consider a sequence of probability measures $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathcal{X})$ such that $\mathcal{W}_2(\mu_n, \mu) \to 0$. Since $\mathcal{X}$ is complete metric space, the sequence $(\mu_n)_{n\in\mathbb{N}}$ is also Cauchy, and therefore, thanks to Proposition 4.2.9 it is tight. Thus, there exists a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ converging narrowly to some $\tilde{\mu} \in \mathscr{P}_2(\mathcal{X})$. Employing Proposition 3.2.7, we have

$$\mathcal{W}_2(\mu, \tilde{\mu}) \le \liminf_{k\to\infty} \mathcal{W}_2(\mu, \mu_{n_k}) = 0.$$

So, we infer that $\mu = \tilde{\mu}$, and thus, the whole sequence goes to μ. Shall we now use a constructive byproduct of the 2-growth condition in order to prove the desired result. More precisely, for any $\epsilon > 0$, one can observe that there exists $C > 0$ such that for every $a, b \ge 0$, it holds that

$$(a+b)^2 \le (1+\epsilon)a^2 + Cb^2. \tag{4.6}$$

Combining (4.6) with classical triangle inequality of d, for every $x_0, x, y \in \mathcal{X}$, we have

$$d^2(x_0, x) \le (1+\epsilon)d^2(x_0, y) + Cd^2(x, y). \tag{4.7}$$

Now, let π_n be the optimal transport plan between μ_n and μ. Integrating (4.7) with respect to the measure π_n and using the fact that π has marginal the measure μ, we have

$$\int_{\mathcal{X}} d^2(x_0, x) d\mu_n(x) \le (1-\epsilon)\int_{\mathcal{X}} d^2(x_0, y) d\mu(y) + x\int_{\mathcal{X}\times\mathcal{X}} d^2(x, y) d\pi_n(x, y).$$

Over and above, by our assumption, we have

$$\int_{\mathcal{X}\times\mathcal{X}} d^2(x, y) d\pi_n(x, y) = \mathcal{W}_2^2(\mu_n, \mu) \to 0, \quad \text{as } n \to \infty,$$

and thusly,

$$\limsup_{n\to\infty} \int_{\mathcal{X}} d^2(x_0, x) d\mu_n(x) \le (1+\epsilon)\int_{\mathcal{X}} d^2(x_0, x) d\mu(x).$$

To conclude, letting $\epsilon \to 0$, we obtain

$$\limsup_{n\to\infty} \int_{\mathcal{X}} d^2(x_0, x) d\mu_n(x) \le \int_{\mathcal{X}} d^2(x_0, x) d\mu,$$

and thanks to Proposition 4.2.7, we have that $\mu_n \to \mu$ in the narrow sense.
Now, let us prove the converse implication $(\Leftarrow)$.
Let us pick an optimal transport plan π_n between μ_n and μ. We know that the sequence of transport plans $(\pi_n)_{n\in\mathbb{N}}$ is tight, so up to subsequences, we assume that it narrow converges to some optimal transport plan π By stability of optimality of optimal transport plans, given in Proposition 4.2.8, we know that the limit optimal transport plan π is still optimal, so we obtain that

$$\int_{\mathcal{X}\times\mathcal{X}} d^2(x,y)d\pi(x,y) = 0.$$

On the other hand, thanks to Proposition 4.2.7, and our assumption on $(\mu_n)_{n\in\mathbb{N}}$ and μ, we know that $(\mu_n)_{n\in\mathbb{N}}$ is 2-uniformly integrable, thus thanks to 2-uniform integrability property which was described in Property 4.2.6, $(\pi_n)_n$ is also 2-uniformly integrable. Now, since the map $(x, y) \mapsto d^2(x, y)$ has quadratic growth in $\mathcal{X} \times \mathcal{X}$, we obtain

$$\lim_{n\to\infty} \mathcal{W}_2^2(\mu_n,\mu) = \lim_{n\to\infty} \int_{X\times\mathcal{X}} d^2(x,y)d\pi_n(x,y) = \int_{\mathcal{X}\times\mathcal{X}} d^2(x,y)d\pi(x,y) = 0,$$

which proves our claim and concludes our proof. □

The above fundamental property of Wasserstein space $\mathbb{W}_2(\mathcal{X})$ lead us to two very important consequence, related with the topological structure of $\mathbb{W}_2(\mathcal{X})$. To be more accurate, on the one hand it provides us with the continuity of $\mathcal{W}_2$, and on the other hand, it shows that if the base space $\mathcal{X}$ is compact, then the Wasserstein space $\mathbb{W}_2(\mathcal{X})$ is also compact. Note, the later fact does not hold in case of locally compactness of the base space $\mathcal{X}$. Both of these facts are presented in the following remarks.

Remark 4.2.11 (Continuity of $\mathcal{W}_2$)**.** A standard corollary, coming from Proposition 4.2.10, is that Wasserstein distance $\mathcal{W}_2$ is *continuous.* Precisely, by a standard application of the triangle inequality, we can see that if $(\mu_n)_{n\in\mathbb{N}}, (\nu_n)_{n\in\mathbb{N}} \in \mathscr{P}_2(\mathcal{X})$, such that $\mu \to \mu$ and $\nu_n \to \nu$, as $n \to \infty$ both in narrow sense, then we have

$$\mathcal{W}_2(\mu_n,\nu_n) \longrightarrow \mathcal{W}_2(\mu,\nu), \qquad \text{as } n\to\infty.$$

Remark 4.2.12 (Compactness versus local compactness in $\mathbb{W}_2(\mathcal{X})$)**.** Another interesting consequence of Proposition4.2.10 is the fact that if $\mathcal{X}$ is *compact,* then the space $\mathbb{W}_2(\mathcal{X})$ is compact too, since the convergence with respect to $\mathcal{W}_2$ is equivalent to narrow convergence.

On the other hand, if $\mathcal{X}$ is unbounded, then the space $\mathbb{W}_2(\mathcal{X})$ is *not* locally compact. In particular, for any probability measure $\mu \in \mathscr{P}_2(\mathcal{X})$ and any $r > 0$, the closed ball of radius equal to r around μ is not compact. We can see this by means of a counterexample. Consider a sequence $(x_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ and let us fix $\overline{x} \in \mathcal{X}$ such that $d(x_n, \overline{x}) \to \infty$. Now let us define the sequence of probability measures $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathcal{X})$ by $\mu_n := (1-\epsilon_n)\mu + \epsilon_n\delta_{x_n}$, where ϵ_n satisfies the relation $\epsilon_n d^2(x_n, \overline{x}) = r^2$. Now, fixing $(1-\epsilon_n)\mu$, moving $\epsilon_n\mu$ to $\overline{x}$ and moving $\epsilon_n\delta_{\overline{x}}$ into $\epsilon x\delta_{x_n}$, we can obtain the bound

$$\mathcal{W}_2^2(\mu_n,\mu) \le \epsilon_n\bigg(\int_{\mathcal{X}} d^2(x_n,\overline{x})d\mu + d^2(x_n,\overline{x})\bigg),$$

and thus,

$$\limsup_{n\to\infty} W_2^2(\mu_n,\mu) \le r.$$

Then, we observe that

$$\liminf_{n\to\infty} \int_{\mathcal{X}} d^2(x,\overline{x})d\mu_n = \liminf_{n\to\infty}(1-\epsilon_n)\int_{\mathcal{X}} d^2(x,\overline{x})d\mu + \epsilon_n d^2(x_n,\overline{x}) = \int_{\mathcal{X}} d^2(x,\overline{x})d\mu + r^2$$

which shows that the 2-moments are not finite. So since the sequence $(\mu_n)_{n\in\mathbb{N}}$ convergences to μ weakly, there is no local compactness.

The results of our investigation of properties Wasserstein spaces, at present, seems to be very much delicated. More precisely, all of the above results have already undercover a nice behaviour of Wasserstein spaces over complete a separable metric spaces. This fact leads us to pose another related question about their topological structure: *is a Wasserstein space over a complete and separable metric space complete and separable metric space too?* Luckily, the answer is positive.

Proposition 4.2.13. *Consider that $(\mathcal{X}, d)$ is a complete and separable metric space. Then the Wasserstein space $\mathbb{W}_2(\mathcal{X}) = (\mathscr{P}_2(\mathcal{X}), \mathcal{W}_2)$ is also complete and separable metric space.*

Proof. Thanks to Proposition 4.2.2, we can infer that $\mathcal{W}_2$ is a metric on $\mathbb{W}_2(\mathcal{X})$, thus, with respect to the metric structure of $\mathbb{W}_2(\mathcal{X})$, there is no concern. It remains to prove that $\mathbb{W}_2(X)$ is separable and complete space. We divide that proof into 2 steps.

Step 1: *Separability*
Since $\mathcal{X}$ is a separable space, there exists a dense sequence $D \subset \mathcal{X}$, and moreover there exists a family $\mathcal{A}$ of probability measures defined as

$$\mathcal{A} = \Big\{\mu \in \mathscr{P}_2(\mathcal{X}) : \mu = \sum_{i=1}^{N} a_i \delta_{x_i} \text{ with } x_i \in D\Big\}.$$

We claim that $\mathcal{A}$ is dense in $\mathscr{P}_2(\mathcal{X})$. To prove such a claim, let $\epsilon > 0$ and $x_0 \in D$. Then if μ belongs to $\mathscr{P}_2(\mathcal{X})$, there exist a compact set $K \subset \mathcal{X}$ such that

$$\int_{\mathcal{X}\setminus K} d^2(x_0, x) d\mu \leq \epsilon^2.$$

Under this consideration, we can cover K by finitely many balls $B(x_i, \epsilon/2)$ for $i = 1, \cdots, n$ with centers $x_i \in D$, and then we can define

$$B'_i = B(x_i, \epsilon)\setminus \bigcup_{j<i} B(x_j, \epsilon).$$

By construction, all of the balls B'_n are disjoint and they still cover the set K. Now, we define the map f on $\mathcal{X}$ by

$$f(B'_n \cap K) = \{x_n\} \quad \text{ and } f(\mathcal{X}\setminus K) = \{x_0\}.$$

Then, for any $x \in K$, we have that $d(x, f(x)) < \epsilon$, and therefore, we can obtain

$$\begin{aligned} \int_{\mathcal{X}} d^2(x, f(x)) d\mu &\leq \epsilon^2 \int_K d\mu + \int_{\mathcal{X}\setminus K} d^2(x_0, x) d\mu && (4.8) \\ &\leq \epsilon^2 + \epsilon^2 = 2\epsilon^2. && (4.9) \end{aligned}$$

Now, since (Id, f) is a transport plan of μ and $f_{\#}\mu$, we have also that

$$\mathcal{W}_2(\mu, f_{\#}\mu) \leq 2\epsilon^2$$

Moreover, $f_{\#}\mu$ can be rewritten as $\sum_{i=1}^{N} a_i \delta_{x_i}$, thus μ can be approximated by a finite combination of Dirac measures. The only thing left to prove is that the coefficients a_i can be replaced by rational coefficients. To prove this, we will employ a total variation bound, the spirit of which was described in Proposition 4.2.4. More precisely, according to total variation bound, we have

$$\mathcal{W}_2\big(\sum_{i=1}^{N} a_i\delta_{x_i}, \sum_{i=1}^{N} b_i\delta_{x_i}\big) \leq 2^{1/2} \cdot \big(\max_{k,\ell\in\mathbb{N}} d(x_k,x_\ell)\big) \sum_{i=1}^{N} |a_i - b_i|^{1/2}.$$

Therefore, since the last term in above expression can be as much small as we want, we can replace coefficients by rational ones, and this fact makes the proof of separability complete.

Step 2: *Completeness*
To prove completeness, we consider a Cauchy sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{P}_2(\mathcal{X})$. Since $\mathcal{X}$ is complete metric space, $(\mu_n)_{n\in\mathbb{N}}$ is convergent, so thanks to Proposition 4.2.9, it is tight. Hence, thanks to Proposition 3.2.4 (Prokhorov's theorem), it admits a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$, which converges to some measure μ. Moreover, we have that

$$\int_{\mathcal{X}} d^2(x_0,x)d\mu \leq \liminf_{k\to\infty} \int_{\mathcal{X}} d^2(x_0,x)d\mu_{n_k} < \infty,$$

and thus, μ belongs to $\mathscr{P}_2(\mathcal{X})$. In addition, thanks to the lower semicontinuity of $\mathcal{W}_2$ (Proposition 4.2.8) and up to a further subsequence $(\mu_{n_{k_\ell}})_{\ell\in\mathbb{N}}$ we have

$$W_2(\mu,\mu_{n_{k_\ell}}) \leq \liminf_{k\to\infty} \mathcal{W}_2(\mu_{n_k},\mu_{n_{k_\ell}})$$

and more precisely,

$$\limsup_{\ell\to\infty} \mathcal{W}_2(\mu,\mu_{n_{k_\ell}}) \leq \limsup_{k,\ell\to\infty} W_2(\mu_{n_k},\mu_{n_{k_\ell}}) = 0.$$

From this fact we infer that the subsequence $(\mu_{n_{k_\ell}})_{\ell\in\mathbb{N}}$ converges to μ in $\mathbb{W}_2(\mathcal{X})$. To conclude, since $(\mu_n)_{n\in\mathbb{N}}$ is Cauchy converging subsequence in $\mathbb{W}_2(\mathcal{X})$, by a standard analytical argument, we infer that it converges in $\mathbb{W}_2(\mathcal{X})$. This fact proves completeness, and moreover, makes our proof complete. □

4.3 Geodesics in Wasserstein spaces

In the previous section, taking under consideration the metric nature and the topological properties of a base space $\mathcal{X}$, we saw that the Wasserstein space $\mathbb{W}_2(\mathcal{X})$ over the corresponding base space, seems to have a very similar and nice behaviour with its base space. Under this perspective, the following question arises: *what about geometrical nature of notions like length, geodesics and convexity in Wasserstein spaces?*

In this section, we will gently present a study of the geometry of Wasserstein spaces. For that reason, we will assume that in what will follows, the base space $(\mathcal{X},d)$ is a complete, separable, metric and geodesic space. Having this assumption as a starting point, we will explore many important geometrical properties of Wasserstein space, which, as we have already mentioned, essentially depends on the geodesic assumptions on its base space.

Let us start this discussion by means of an example. We recall that if $x \in \mathcal{X}$, the map $x \mapsto \delta_x \in \mathscr{P}_2(\mathcal{X})$ is an isometry. Therefore, if $t \mapsto \pi_t$ is a constant speed geodesic connecting x and y, the curve $t \mapsto \delta_{\pi_t}$ is a constant speed geodesic on $\mathscr{P}_2(\mathcal{X})$ which connects δ_x to δ_y. The important thing to notice here is that the natural way to interpolate between δ_x and δ_y is given by this -so called- *displacement interpolation* or *McCann's interpolation* (see McCann (1997)), a very important result which came in the mid 90s. Conversely, let us observe that the classical linear interpolation

$$t \mapsto \mu_t := (1-t)\delta_x + t\delta_y$$

generates a curve with infinite length(!) as soon as $x \neq y$, (since $\mathcal{W}_2(\mu_s, \mu_t) = |t-s|^{1/2} d(x,y)$), and for this reason is unnatural to work with this in such a setting.

But, could we have a desirable characterization of constant speed geodesics in Wasserstein space in terms of transport plans?. Favourably, the answer is positive, but in order to prove such a result, we need the following lemma, which stands as a closed graph argument. Its proof is omitted since is constructed by straightforward arguments.

Lemma 4.3.1. *Consider that $(\mathcal{X}, d)$ is a complete and separable geodesic metric space. Then the multivalued map $G : \mathcal{X} \times \mathcal{X} \to \mathscr{G}(\mathcal{X})$ which associates each pair (x, y) to the set $G(x, y)$ of the constant speed geodesic connecting x and y has closed graph.*

Proposition 4.3.2. *Consider that $(\mathcal{X}, d)$ is complete and separable geodesic metric space. Then the following are equivalent:*

(i) the map $t \mapsto \mu_t \in \mathscr{P}_2(\mathcal{X})$ is a constant speed geodesic

(ii) There exists a measure $\boldsymbol{\mu} \in \mathscr{P}_2(\mathscr{G}(\mathcal{X}))$ such that the measure $(e_0, e_1)_\# \boldsymbol{\mu}$ is a optimal transport plan between μ_0 and μ_1 and

$$\mu_t = (e_t)_\# \boldsymbol{\mu}.$$

Proof. This proof is highly constructive and combines several important results. We will present a skerch of it.
$(ii) \Rightarrow (i)$: We choose $\mu^{(0)}, \mu^{(1)} \in \mathscr{P}_2(\mathcal{X})$ and we find an optimal transport plan γ between them. By Lemma 4.3.1 and thanks to classical measurable selection theorems (see for details Aliprantis & Border (2006) and Kechris (2012)), we know that there exists a Borel map $G_s : \mathcal{X} \times \mathcal{X} \to \mathscr{G}(\mathcal{X})$ such that for every $x, y \in \mathcal{X}$ the curve $G_s(x, y)$ is a constant speed geodesic connecting x to y. We define the Borel probability measure $\boldsymbol{\mu} \in \mathscr{P}(\mathscr{G}(\mathcal{X}))$ by

$$\boldsymbol{\mu} := G_{s\#}\gamma,$$

and, using the evaluation maps, we define the measures $\mu_t \in \mathscr{P}(\mathcal{X})$ by $\mu_t := (e_t)_\# \boldsymbol{\mu}$.

Our claim is that the map $t \mapsto \mu_t$ is a constant speed geodesic connecting $\mu^{(0)}$ to $\mu^{(1)}$. To see this, we consider the map $(e_0, e_1) : \mathscr{G}(\mathcal{X}) \to \mathcal{X} \times \mathcal{X}$ and we observe that since $(e_0, e_1)\big(G_s(x, y)\big) = (x, y)$, we obtain

$$(e_0, e_1)_\# \boldsymbol{\mu} = \gamma.$$

Moreover, we have that $\mu_0 = (e_0)_\# \boldsymbol{\mu} = (\mathsf{p}_1)_\# \gamma = \mu^{(0)}$, and in the same fashion, $\mu_1 = \mu^{(1)}$. Thus, the curve $t \mapsto \mu_t$ connects $\mu^{(0)}$ to $\mu^{(1)}$. In addition, for $s < t$ we have that

$$\begin{aligned}
\mathcal{W}_2^2(\mu_t, \mu_s) &\leq \int_{\mathcal{X}\times\mathcal{X}} d^2\big(e_t(\gamma) e_s(\gamma)\big) d\boldsymbol{\mu}(\gamma) \\
&= (t-s)^2 \int_{\mathcal{X}\times\mathcal{X}} d^2\big(e_0(\gamma), e_1(\gamma)\big) d\boldsymbol{\mu}(\gamma) \\
&= (t-s)^2 \int_{\mathcal{X}\times\mathcal{X}} d^2(x, y) d\gamma(x, y) \\
&= (t-s)^2 \mathcal{W}_2^2(\mu^{(0)}, \mu^{(1)}),
\end{aligned}$$

which proves that the measures μ_t have finite second moment and (μ_t) is a constant speed geodesic. Therefore, our claim is proved.
$(i) \Rightarrow (ii)$: For $n \geq 0$, we use iteratively the Gluing lemma given by Proposition 4.2.1 and the Borel map $G_s : \mathcal{X} \times \mathcal{X} \to \mathscr{G}(\mathcal{X})$ to construct a measures $\boldsymbol{\mu}^n \in \mathscr{P}\big(C([0,1]), \mathcal{X})\big)$ such that

$$(e_{i/2^n}, (e_{(i+1)/2^n})_\# \boldsymbol{\mu}^n$$

is an optimal transport plan between $\mu_{i/2^n}$ and $\mu_{(i+1)/2^n}$, for every $i = 0, \cdots 2^n - 1$, and $\boldsymbol{\mu}^n$-a.e. γ be a geodesic in intervals $[i/2^n, (i+1)/2^n]$, for $i = 0, \cdots 2^n - 1$. Then, we fix a $n \geq 0$ and observe that for any $0 \leq j \leq k \leq 2^n$ it holds that

$$\begin{aligned} \|d(e_{j/2^n}, e_{k/2^n})\|_{L^2(\boldsymbol{\mu}^n)} &\leq \left\| \sum_{i=j}^{k-1} d(e_{i/2^n}, e_{(i+1)/2^n}) \right\|_{L^2(\boldsymbol{\mu}^n)} \leq \sum_{i=j}^{k-1} \|d(e_{i/2^n}, e_{(i+1)/2^n})\|_{L^2(\boldsymbol{\mu}^n)} \quad (4.10) \\ &= \sum_{i=j}^{k-1} \mathcal{W}_2(\mu_{i/2^n}, \mu_{(i+1)/2^n} = \mathcal{W}_2(\mu_{j/2^n}, \mu_{k/2^n}). \quad (4.11) \end{aligned}$$

Thus, we have that $(e_{j/2^n}, e_{k/2^n})_{\#}\boldsymbol{\mu}^n$ is an optimal transport plan between $\mu_{j/2^n}$ and $\mu_{k/2^n}$, for $0 \leq j \leq k \leq 2^n$. Note also that since the inequalities given by (4.10) are equalities, we infer that for $\boldsymbol{\mu}^n$-a.e. γ the points $\gamma_{i/2^n}$, for $i = 0, \cdots, 2^n$, must lie along a geodesic and satisfy

$$d(\gamma_{i/2^n}, \gamma_{(i+1)/2^n} = d(\gamma_0, \gamma_1)/2^n), \text{ for } i = 0, \cdots, 2^n - 1.$$

Hence, γ is a constant speed geodesic $\boldsymbol{\mu}^n$-a.e. and thus $\boldsymbol{\mu}^n \in \mathscr{P}(\mathscr{G}(\mathcal{X}))$. We assume now that $\boldsymbol{mu}^n$ converges in narrow sense, up to a subsequence, to some $\boldsymbol{\mu} \in \mathscr{P}(\mathscr{G}(\mathcal{X}))$. Then, by the continuity of evaluation maps e_t we have that for any $t \in [0,1]$, the sequence $\big((e_t)_{\#}\boldsymbol{\mu}^n\big)_n$ converges in narrow sense to $(e_t)_{\#}\boldsymbol{\mu}$. Combining this fact with the uniform bound which was given by (4.3), we have that

$$\mu_t = (e_t)_{\#}\mu.$$

Therefore, now we have to prove that some subsequence has a narrow limit. We will prove it by showing that $\boldsymbol{\mu}^n \in \mathscr{P}_2(\mathscr{G}(X))$ for every $n \in \mathbb{N}$ and that subsequence is Cauchy sequence in $\mathbb{W}_2(\mathscr{G}(\mathcal{X}))$, and then thanks to Proposition 4.2.7, we obtain the desired result.

Thanks to Theorem 3.2.4 (Prokorov's theorem), we know that the set of transport plans in tight. Moreover, thanks to 2-uniform integrability and convergence properties in Wasserstein spaces, we have that for every $n \in \mathbb{N}$ the set of plans $\alpha \in \mathscr{P}_2(\mathcal{X}^{2n+1})$ such that

$$(\mathsf{p}_i)_{\#}\alpha = \mu_{i/2_n}, \qquad \text{for } i = 0, \cdots, 2^n$$

is compact in $\mathscr{P}_2(\mathcal{X}^{2n+1})$. Thus, employing a diagonal argument, we pass to a subsequence, and then assume that for every $n \in \mathbb{N}$ the sequence

$$\left(\prod_{i=0}^{2n} (e_{i/2^n})_{\#}\boldsymbol{\mu}^m \right)_{m \in \mathbb{N}}$$

converges to some transport plan with respect to $\mathcal{W}_2$ defined on $\mathcal{X}^{2^n+1}$. Fixing now $n \in \mathbb{N}$, we notice that for $t \in [i/2^n, (i+1)/2^n]$ and $\gamma, \tilde{\gamma} \in \mathscr{G}(\mathcal{X})$ it holds that

$$d(\gamma_t, \tilde{\gamma}_t) \leq d(\gamma_{i/2^n}, \tilde{\gamma}_{(i+1)/2^n}) + \frac{1}{2^n}\big(d(\gamma_0, \gamma_1) + d(\tilde{\gamma}_0, \tilde{\gamma}_1)\big).$$

So, squaring and taking supremum over $t \in [0,1]$, we obtain

$$\sup_{t \in [0,1]} d^2(\gamma_t, \tilde{\gamma}_t) \leq 2 \sum_{i=0}^{2^n-1} d^2(\gamma_{i/2^n}, \tilde{\gamma}_{(i+1)/2^n}) + \frac{1}{2^{n-2}}\big(d^2(\gamma_0, \gamma_1) + d^2(\tilde{\gamma}, \tilde{\gamma}_1)\big). \tag{4.12}$$

Now, choosing $\tilde{\gamma}$ to be a constant speed geodesic and using the estimate (4.3), we get that $\boldsymbol{\mu}^m \in \mathscr{P}_2(\mathscr{G}(\mathcal{X}))$ for every $\mu \in \mathbb{N}$. Then, for any given $\nu, \tilde{\nu} \in \mathscr{P}(\mathscr{G}(\mathcal{X}))$, due to Gluing lemma (Lemma 4.2.1), for $\mathcal{Y} = \mathscr{G}(\mathcal{X})$ and $\mathcal{Z} = \mathcal{X}^{2n+1}$ we can find a plan $\boldsymbol{\beta} \in \mathscr{P}(\mathscr{G}(\mathcal{X}) \times \mathscr{G}(\mathcal{X}))$ such that

$$(\mathsf{p}_1)_{\#}\boldsymbol{\beta} = \nu \quad \text{ and } (\mathsf{p}_2)_{\#}\boldsymbol{\beta} = \tilde{\nu}$$

and the measure

$$\Big(\big(e_0, \cdots, e_{i/2^n}, \cdots, e_1\big) \circ \big(e_0, \cdots, e_{i/2^n}, \cdots, e_1\big) \circ \mathsf{p}_2 \Big)_{\#} \boldsymbol{\beta}$$

is a optimal transport plan between $\prod_{i=0}^{2^n}(e_{i/2^n})_\#\nu$ and $\prod_{i=0}^{2^n}(e_{i/2^n})_\#\tilde{\nu}$, where optimality stands with respect to $\mathcal{W}_2$ defined on $\mathscr{P}_2(\mathcal{X}^{2n+1})$. Now, using $\boldsymbol{\beta}$ to bound from above the distance $\mathcal{W}_2(\nu, \tilde{\nu})$ and thanks to the bound given by (4.12), we obtain that for every pair of measures $\nu, \tilde{\nu} \in \mathscr{P}_2(\mathscr{G}(\mathcal{X}))$ it holds

$$\mathcal{W}_2^2(\nu, \tilde{\nu}) \leq 2\mathcal{W}_2^2\bigg(\prod_{i=0}^{2^n}(e_{i/2^n})_\#\nu, \prod_{i=0}^{2^n}(e_{i/2^n})_\#\tilde{\nu} \bigg) + \frac{1}{2^n - 2}\bigg(\int_{\mathcal{X}\times\mathcal{X}} d^2(\gamma_0, \gamma_1) d\nu(\gamma) + \int_{\mathcal{X}\times\mathcal{X}} d^2(\tilde{\gamma}_0, \tilde{\gamma}_1) d\nu(\tilde{\gamma}) \bigg).$$

Plugging $\nu = \boldsymbol{\mu}^m$ and $\tilde{\nu} = \boldsymbol{\mu}^m$ and recalling that

$$\mathcal{W}_2^2\bigg(\prod_{i=0}^{2^n}(e_{i/2^n})_\#\boldsymbol{\mu}^m, \prod_{i=0}^{2^n}(e_{i/2^n})_\#\boldsymbol{\mu}^{m'} \bigg) \longrightarrow 0, \qquad \text{as } m, m' \to \infty.$$

Thus, for every $n \in \mathbb{N}$, we get

$$\begin{aligned} \limsup_{m,m'\to\infty} \mathcal{W}_2(\boldsymbol{\mu}^m, \boldsymbol{\mu}^{m'}) &\leq \frac{1}{2^n - 2}\bigg(\int_{\mathcal{X}\times\mathcal{X}} d^2(\gamma_0, \gamma_1) d\boldsymbol{\mu}^m(\gamma) + \int_{\mathcal{X}\times\mathcal{X}} d^2(\tilde{\gamma}_0, \tilde{\gamma}_1) d\boldsymbol{\mu}^m(\tilde{\gamma}) \bigg) \\ &= \frac{1}{2^{n-3}} \mathcal{W}_2(\mu_0, \mu_1). \end{aligned}$$

Letting n go to infinity, we obtain that $(\boldsymbol{\mu}^m) \subset \mathscr{P}(\mathscr{G}(\mathcal{X}))$ is Cauchy, and the desired result follows. □

So, since now we have and general and quite abstract characterization of constant speed geodesics in Wasserstein spaces. One might ask how this characterization can be reduced, in classical base space settings, like Hilbert spaces. Reducing such a result to Hilbert space setting, we can infer several nice phenomena, which are presented in the following remark.

Remark 4.3.3 (Constant speed geodesics in Hilbert space as base space)**.** Consider that $\mathcal{X}$ is a Hilbert space. Then for every $x, y \in \mathcal{X}$ there exists *only* one constant speed geodesic connecting x to y, that is the curve $t \mapsto (1-t)x + ty$. Therefore, Proposition 4.3.2 reads as follows: the map $t \mapsto \mu_t$ is a constant speed geodesic if and only if there exists an optimal transport plan γ between μ_0 and μ_1 such that

$$\mu_t = \big((1-t)\mathsf{p}_1 + \mathsf{p}_2\big)_\#\gamma. \tag{4.13}$$

If also γ is induced by a transport map $\boldsymbol{T}$, the equation (4.13) becomes the classical interpolation

$$\mu_t = \big((1-t)\mathsf{Id} + t\boldsymbol{T}\big)_\#\mu_0.$$

The above proof, except of its elegant consequences in simpler settings, such Hilbert spaces, make us able to prove a very desired result. Accurately, based on such constructions, we can prove that Wasserstein spaces enjoy several properties related to geometric features of their base space. Precisely, we have the following result.

Theorem 4.3.4 (If $\mathcal{X}$ is a geodesic space, $\mathbb{W}_2(\mathcal{X})$ is too.)**.** *Let $(\mathcal{X}, d)$ be a complete separable and geodesic metric space. Then the space $\mathbb{W}_2(\mathcal{X})$ is also geodesic space.*

Proof. Since we have done the dirty work in the second part of Proposition 4.3.2, it is possible to infer that if $x, y \in \mathcal{X}$ and a geodesic $(\mu_t)_{t\in[0,1]}$ connecting δ_x to δ_y, using very similar arguments we can construct a measure $\mu \in \mathscr{P}(\mathscr{G}(\mathcal{X}))$ such that

$$\mu_t = (e_t)_\#\mu.$$

Then, every $\gamma \in \operatorname{supp}(\mu)$ is a geodesic connecting x to y, and the desired result follows. □

4.4 Dynamical formulation and Riemannian structure

At this point, let us observe one important fact of optimal transport problems. In the previous chapter, the formulation of optimal transport problem was developed in *static* framework. This means that, under our previous perspective, the whole transportation process was depending only on the optimal transport plan. Once we have fully understood the optimal transport plan, the only one to have to do someone is to move points from initial position to their final one through optimal transport map. But, what about the *dynamic formulation* of optimal transport problems? Perhaps, all of these investigations have as starting point the seminal paper of J. D. Benamou and Y. Brenier (see Benamou & Brenier (2000)), and nowadays consists a quite active research field.

To make the discussion more concrete, as is widely know in several situations with dynamic modelling issues (like fluid mechanics), it is classical to consider two different but complementary ways to study motion phenomena, that is *Lagrangian* and *Eulerian.*

Under the Lagrangian point of view, we see the motion of cloud of particles (or some fluid) as a parcel, which moves in space and time. Then, we study at every time frame and every parcel, what happens to that parcel. In other words, given two different times, $t = 0$ and $t = 1$, the motion of a particle from a position x to a position y is characterized by its initial and final position. Under this perspective, this descriptions resembles to the optimal transport problem, where the *transport plan* is nothing more but a probability measure on these pairs (x, y). Thusly, the form of optimal transport problem, which we have already seen, is stated in Lagrangian spirit.

On the contrary, the Eulerian point of view, studying the behaviour of a cloud of particles (of some fluid), we are interested about what happens at every time t and every location x, describing quantities like velocity, density of rate of flow at each time and each point. So, in this context, the dynamical modelling of a cloud of particles (or some fluid) could be started by considering the density of them $\varrho(t, x)$ and the velocity $v_t(x)$ of each of particle at time t. In this spirit, one can write the equations which are satisfied by the density of the cloud of particles according to the velocity field v. This task, that we prescribe the initial density ϱ_0 and that the position of particle initially originated at point x will be obtained by the solution of the following ordinary differential equation:

$$\begin{aligned} y_x'(t) &= v_t(y_x(t)) \\ y_x(0) &= x. \end{aligned}$$

We can define the flow map $Y_t(x) := y_x(t)$ and then look for a measure $\varrho_t := (Y_t)_{\#}\varrho_0$. Under this persepective, we expect that the pair (ϱ_t, v_t) will solve the continuity equation, that is

$$\frac{d}{dt}\varrho_t + \nabla \cdot (\varrho_t v_t) = 0.$$

So, under these considerations, we can understand that the continuity equation describes the link between the evolution of density ϱ_t and the instantaneous velocity v_t of every point of x. With this fact in our mind, it is natural to think the velocity field v_t as the infinitesimal variation of continuum ϱ_t.

We shall now present the connection between the above discussion and Wasserstein spaces. For that reason, we consider that $\mathcal{X} = \mathbb{R}^d$ and $(\mu_t)_{t\in[0,1]} \subset \mathscr{P}_2(\mathbb{R}^d)$, is a constant speed geodesic induced by some optimal transport map $\boldsymbol{T}$, which means that it has the form

$$\mu_t := \big((1-t)\mathsf{Id} + t\boldsymbol{T}\big)_{\#}\mu_0.$$

Now, taking a test function $\phi \in C_c^\infty(\mathbb{R}^d)$ and employing integration by parts, we can arrive to

$$\frac{d}{dt}\int_{\mathcal{X}} \phi d\mu_t = \frac{d}{dt}\int_{\mathcal{X}} \phi\big((1-t)\mathsf{Id} + t\boldsymbol{T}\big)d\mu_0 = \int_{\mathcal{X}} \langle \nabla\phi\big((1-t)\mathsf{Id} + t\boldsymbol{T}\big), \boldsymbol{T} - \mathsf{Id}\rangle d\mu_0 = \int_{\mathcal{X}} \langle \nabla\phi, v_t\rangle d\mu_t,$$

which means that $(\mu_t)_{t\in[0,1]}$ satisfies the continuity equation

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t\mu_t) = 0,$$

with $v_t := (\boldsymbol{T} - \mathsf{Id}) \circ ((1-t)\mathsf{Id} + t\boldsymbol{T})^{-1}$ for every $t \in [0,1]$, in sense of distributions.

From this point of view, one might expect that the curves in Wasserstein space $\mathbb{W}_2(\mathbb{R}^d)$ (or perhaps with a different base space like a separable Hilbert space or a Riemannian manifold), are somehow linked with the solution of the continuity equation. Luckily, this expectation is true, and moreover, this connection can provide us a generic characterization of absolutely continuous curves in Wasserstein space, which is has been done by L. Ambrosio, N. Gigli and G. Savaré (see L. Ambrosio et al. (2008)) and independently by S. Lisini (see Lisini (2007)) employing different arguments. This characterization is presented in the following proposition. Note that, for sake of simplicity, we consider as a base space the Euclidean one, following the presentation of the survey of Brasco (2012).

Proposition 4.4.1 (Ambrosio, Gigli, Savaré). *Consider a curve $(\mu_t)_{\in[0,1]} \subset \mathbb{W}_2(\mathbb{R}^d)$ satisfying the continuity equation*

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t\mu_t) = 0 \quad \text{in } \mathbb{R}^d \times (0,1)$$

in sense of distributions, for some Borel vector field $v_t : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ with $\|v_t\|_{L^2(\mu_t)} \in L^2(\mu_t)$, then $(\mu_t)_{t\in[0,1]}$ is an absolutely continuous curve and

$$|\mu'|(t) \leq \|v_t\|_{L^2(\mu_t)}.$$

Conversely, if $(\mu_t)_{t\in[0,1]}$ is a absolutely continuous curve with $|\mu'| \in L^2(\mu_t)$ be its metric derivative. Then there exists Borel vector fields $v_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ such that $v_t \in L^2(\mu_t)$ with $\|v_t\|_{L^2(\mu_t)} \leq |\mu'|(t)$ for a.e. $t \in [0,1]$.

Proof. ($\Rightarrow$) Let us start heuristically, by observing the following fact: since there is not regularity assumption of the curve μ_t, based on the *method of characteristics* (see L. Ambrosio et al. (2008), Chapter 8), it would be of the form $(X_t)_{\#}\mu_0$, where X_t is the flow map of v_t, that is

$$\begin{aligned} X_t'(x) &= v_t(X_t(x)) \\ X_0(x) &= x. \end{aligned}$$

The, fixing $s, t \in [0,1]$, we can estimate the quantity $\mathcal{W}_2(\mu_s, \mu_t)$, using the transport plan $\pi_{st} := (X_s \times X_t)_{\#}\mu_0$ by

$$\mathcal{W}_2^2(\mu_s, \mu_t) \leq \int_{\mathbb{R}^d} |X_s(x) - X_t(x)|^2 d\mu_0(x).$$

Moreover, we observe that

$$X_t(x) - X_s(x) = \int_s^t X_r' dr = \int_s^t v_r\big(X_r(x)\big)dr,$$

and using Cauchy-Schwartz and Jensen inequalities, we arrive to

$$W_2^2(\mu_t, \mu_s) \leq |t-s| \int_{\mathbb{R}^d} \int_s^t |v_r\big(X_r(s)\big)|^2 dr d\mu_0(x). \tag{4.14}$$

At this point, *if* we could interchanging the above integrals (which we actually can) and using the definition of push-forward measure, we could obtain

$$\frac{\mathcal{W}_2^2(\mu_t, \mu_s)}{|t-s|^2} \leq \frac{1}{|s-t|} \int_s^t \int_{\mathbb{R}^d} |v_r(x)|^2 d\mu_r(x) dr.$$

Then taking the limit as $s \to t$, we would obtain the desired result. So far, this was a very heuristic argument.

In general, we do not hav the Lipschitz property, so we have to restart from a regularization argument. To do this, we consider $\mu_t^\epsilon := \mu_t * \rho_\epsilon$ and $\phi_t^\epsilon := (v_t\mu_t) * \rho_\epsilon$, where ρ_ϵ is the convolution kernel supported in the whole $\mathbb{R}^d$. Then, we can see that μ_t^ϵ solves the continuity equation with the smooth velocity field v_t^ϵ implicitly defined by $\phi_t^\epsilon := v_t^\epsilon \mu_t^\epsilon$. Under this perspective, we have that $\mu_t^\epsilon = (X_t^\epsilon)_\# \mu_0^\epsilon$ where X_t^ϵ is the flow map of v_t^ϵ and all of above calculations are justified.

Then, rewriting the right-hand side of (4.14), we obtain

$$\int_s^t \int_{\mathbb{R}^d} |v_t^\epsilon|^2 d\mu_t^\epsilon dr = \int_s^t \int_{\mathbb{R}^d} \Big| \frac{\phi_r^\epsilon(x)}{\phi_r^\epsilon(x)} \Big| \mu_r^\epsilon(x) dx dr.$$

In addition, we observe that the map $(\mu, \phi) \mapsto |\phi|\mu$ is convex, both for μ and ϕ, so again employing a Jensen-inequality argument, we obtain

$$\int_{\mathbb{R}^d} \Big| \frac{\phi_r^\epsilon(x)}{\mu_r^\epsilon(x)} \Big|^2 \mu_r^\epsilon(x) dx \leq \int_{\mathbb{R}^d} \Big| \frac{d\phi_r}{d\mu_r}(x) \Big|^2 d\mu_r(x)$$

From this fact, we conclude that

$$\frac{\mathcal{W}_2^2(\mu_s^\epsilon, \mu_t^\epsilon)}{|t-s|^2} \leq \frac{1}{|t-s|} \int_{[s,t]\times\mathbb{R}^d} \Big| \frac{d\phi_r}{d\mu_r}(x) \Big|^2 d\mu_r(x) dr = \frac{1}{|t-s|} \int_s^t \int_{\mathbb{R}^d} |v_r(x)|^2 d\mu_r(x) dr.$$

To conclude, using the lower semicontinuity property of $\mathcal{W}_2$ (Proposition 4.2.8) and letting $\epsilon \to 0$, the desired result follows.

($\Leftarrow$) To prove this part is a bit more tricky. A detailed treatment of its can be found in L. Ambrosio et al. (2008). We will present a sketch of a constructive idea which was presented in Lisini (2007). Given an absolutely continuous curve $(\mu_t)_{t\in[0,1]} \subset \mathbb{W}_2(\mathbb{R}^d)$, we can always find a probability measure $Q \subset \mathscr{P}_2(AC^2((0,1);\mathbb{R}^d)$ such that $\mu_t = (e_t)_\# Q$, where e_t is the evaluation map at time t, satisfying $e_t(\sigma) = \sigma(t)$ for any continuous curve $\sigma : [0,1] \to \mathbb{R}^d$. Intuitively, this means that we can understand absolutely continuous curves in $\mathbb{W}_2(\mathbb{R}^d)$ as a superposition of absolutely continuous curves of their base space $\mathbb{R}^d$.

In particular, these curves could be chosen is such a way, that we can control their 2-moment velocities with their metric derivatives with respect to $\mathcal{W}_2$, or more precisely, we can construct a probability measure $Q \subset \mathscr{P}_2(AC^2((0,1);\mathbb{R}^d)$ which satisfies

$$\left(\int_{AC^2((0,1);\mathbb{R}^d)} |\sigma'(t)|^2 dQ(\sigma) \right)^{1/2} \leq |\mu_t'|, \quad \text{for } t \in [0,1]. \tag{4.15}$$

The proof of such a claim is quite long and technical. Nevertheless, the underlying idea is quite concrete and its briefly described in the following.

We consider a partition $0 = t_0 \leq t_1 \leq \cdots \leq t_{k+1} \leq 1$ of $[0,1]$ and we discretize the curve $(\mu_t)_{t\in[0,1]}$ by considering the measures $\{\mu_{t_0}, \cdots, \mu_{t_{k+1}}\}$. For any $0 \leq i \leq k+1$, we interpolate between μ_{t_i} and $\mu_{t_{i+1}}$ using an optimal transport plan $\gamma_{i,i+1}$. By this procedure, we construct a measure which is concentrated on linear curves (the so-called *transport rays*) and parametrized on $[t_i, t_{i+1}]$. Then, "gluing" together all of these measures, we end up to a probability measure $Q_k \subset \mathscr{P}_2(AC^2((0,1);\mathbb{R}^d)$, which is concentrated on piecewise linear curves and satisfying (4.15). Moreover, by choosing a dyadic partition and taking the limit as $k \to \infty$, we can proof a tightness-type property of the sequence $(Q_k)_{k\in\mathbb{N}}$.

So, once we have this construction in our toolbox, we can consider the disintegration of Q with respect to evaluation map e_t, concluding to the fact that

$$Q = \int_0^1 Q_x^t d\mu_t,$$

where Q_x^t stands for a Borel probability measure concentrated on the fiber $e_t^{-1}(\{x\}) = \{\sigma : \sigma(t) = x\}$. Then, we can construct the desired vector field v_t as the average of velocities of curves corresponding to Q, that is

$$v_t(x) := \int_{\{\sigma:\sigma(t)=x\}} \sigma'(x) dQ(\sigma).$$

Under this perspective, we can show that (μ_t, v_t) solves the continuity equation, and thanks to (4.15), we have that $v_t \in L^p(\mathbb{R}^d; \mu_t)$ and $\|v_t\|_{L^p(\mathbb{R}^d;\mu_t)} \leq |\mu_t'|$, which makes the sketch of our proof complete. □

The above result can be (almost) easily extended to any separable Hilbert space (see L. Ambrosio et al. (2008)), and using Nash's emdedding theorem, it can be (alsmost) tough extended to a Riemannian manifold setting (see Ambrosio L., Gigli N. (2013) for a skerch of proof).

Moreover, the above characterization is very important for many reasons and it has consequences in many different situations. Let us mention three of them.

Firstly, resembling the Riemannian manifold setting, the above characterization of curves in Wasserstein space through continuity equation strongly suggest that the scalar product in $L^2(\mu)$ should essentially be considered as the *metric tensor* in the Wasserstein space over that Riemmanian manifold at point μ. Secondly, since there is no-regularity assumption on $(\mu_t)_{t\in[0,1]}$, the link of continuity equation seems to be very powerful in order to treat several PDEs, by taking the following program: translate a curve in Wasserstein space into a PDE and then employ modern PDE methods, like viscosity solutions (see e.g. Gangbo et al. (2008)), to study them. Thirdly, based on this characterization, one can (almost) directly obtain, the famous *Benamou-Brenier* formula, as a standard corollary[3] Following the latter reasoning, we present the Benamou-Brenier formula in the following.

Corollary 4.4.2 (Benamou-Brenier). *Consider two probability measures $\nu, \nu' \in \mathbb{W}_2(\mathbb{R}^d)$. Then it holds that*

$$\mathcal{W}_2(\nu, \nu') = \inf \left\{ \int_0^1 \|v_t\|_{L^2(\mu_t)} dt \right\}, \tag{4.16}$$

where the infimum is taken over the distributional solutions of the continuity equation (μ_t, v_t) such that $\nu = \mu_0$ and $\nu' = \mu_1$

Proof. Let (μ_t, v_t) be a solution of continuity equation. If $\int_0^1 \|v_t\|_{L^2(\mu_t)} = +\infty$ and proposition becomes trivial.

In other cases, we apply the first part of Proposition 4.4.1 to obtain an absolutely continuous curve $(\mu_t)_{t\in[0,1]} \subset \mathscr{P}_2(\mathbb{R}^d)$, and thus, we can arrive to

$$\mathcal{W}_2(\nu, \nu') \leq \int_0^1 |\mu_t'| dt \leq \int_0^1 \|v_t\|_{L^2(\mu_t)} dt.$$

Therefore, our proof will be complete if we prove the converse inequality. To do this, it suffices to consider a constant speed geodesic $(\mu_t)_{t\in[0,1]}$ connecting ν to ν', and applying the first part of

[3] We have to emphasize that Benamou-Brenier formula was a pre-existing result than Ambrosio, Gigli, Savaré's characterization. Although, it can be achieved as chic corollary using the characterization of absolute continuous curves through continuity equation.

Proposition 4.4.1. From this, we obtain the existence of vector fields v_t such that the continuity equations is satisfied and

$$\|v_t\|_{L^2(\mu_t)} \leq |\mu_t'| = \mathcal{W}_2(\nu, \nu'), \quad \text{a.e.} \quad t \in [0,1].$$

Thus, we have that

$$\mathcal{W}_2(\mu_0, \mu_1) \geq \int_0^1 \|v_t\|_{L^2(\mu_t)} dt. \tag{4.17}$$

Combining (4.16) and (4.17), the desired result follows. □

Based on the dynamical formulation of optimal transport problem, let us turn now our attention to another important phenomenon. According to above discussion, one can also observe that, in general, given an absolutely continuous curve $(\mu_t)_{t\in[0,1]} \in \mathbb{W}_2(\mathbb{R}^d)$, there is no unique choice of vector field (v_t) such that the continuity equation is satisfied. Such a non-uniqueness can be easily seen by the fact that if the continuity equation holds true and (w_t) is a Borel family of vector fields such that

$$\nabla \cdot (w_t, \mu_t) = 0, \quad \text{a.e. } t \in [0,1].$$

then the continuity equation is *also* satisfied with vector fields $(v_t + w_t)_t$. A natural question arising from this observation is whether there exists a "way", in some sense, to associate uniquely a family of vector fields to a given absolutely continuous curve. To answer such a question, we can follow two, different but equivalent strategies.

The first strategy, having an algebraic essence, traces the roots from the fact that the distributional solutions of the continuity equation, act only on the gradient of smooth functions and suggests that the family of v_t should be taken in the set of gradients. Making this fact more rigorous, we can see that v_t should live in

$$\overline{\{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu_t)} \quad \text{for a.e.} \quad t \in [0,1].$$

The second strategy, has a variational essence, and is based on the observation that the continuity equation is linear in v_t and the L^2-norm is strictly convex. This fact implies that there exists a unique family of vector fields $v_t \in L^2(\mu_t)$, for a.e. $t \in [0,1]$, among the vector fields compatible with the absolutely continuous curve μ_t via the continuity equation. In other words, for any other vector field $\tilde{v}_t$, compatible with the curve μ_t in sense the that the continuity equation is satisfied, it holds that

$$\|\tilde{v}_t\|_{L^2(\mu_t)} \geq \|v_t\|_{L^2(\mu_t)}, \quad \text{for a.e. } t \in [0,1].$$

Then, it is not hard to verify that v_t has the minimal norm if and only if lives in the set

$$\left\{ v \in L^2(\mu_t) : \int \langle v, w \rangle d\mu_t = 0 : \text{ for any } w \in L^2(\mu_t) \text{ with } \nabla \cdot (w\mu_t) = 0 \right\}.$$

These two strategies motivate the definition of *tangent space* for some probability measure in $\mathscr{P}_2(\mathbb{R}^d)$.

Definition 4.4.3 (Tangent space)**.** Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. We define the tangent space $\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ of $\mathscr{P}_2(\mathbb{R}^d)$ in μ as

$$\begin{aligned} \mathrm{Tan}_\mu \mathscr{P}_2((\mathbb{R}^d) &:= \overline{\{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu_t)} \\ &= \left\{ v \in L^2(\mu_t) : \int_{\mathbb{R}^d} \langle v, w \rangle d\mu_t = 0 : \text{ for any } w \in L^2(\mu_t) \quad \text{with } \nabla \cdot (w\mu_t) = 0 \right\}. \end{aligned}$$

Based on this definition of the tangent space of $\mathscr{P}_2(\mathbb{R}^d)$, and since it is defined as a closed subset of $L^2(\mu_t)$, we can see that it is endowed with a scalar product on $L^2(\mu)$. Although, working on the space $\mathbb{W}_2(\mathbb{R}^d)$, we can realize that we cannot define an exponential map from a neighbourhood of $\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ into $\mathscr{P}_2(\mathbb{R}^d)$ which is an homeomorphism, and therefore, the space $\mathbb{W}_2(\mathbb{R}^d)$, is *not* a Riemannian manifold. All of these facts resembles a Riemannian flavor of Wasserstein spaces. So, standing on geometric features and analogies with Riemannian case (see Carrillo et al. (2006)), we could say that the characterization of curves given in Proposition 4.4.1, is the core component of the so-called *Riemannian structure* of $\mathbb{W}_2(\mathbb{R}^d)$. Perhaps, the most appropriate analogy is that the Wasserstein metric is the Riemannian metric induced by the tangent bundle, and its inner product is defined as above. This fact was noticed at first by J.D. Benamou and Y. Brenier (see Benamou & Brenier (2000)) in Euclidean space setting and its related with the Benamou-Brenier formula, which we have already discussed. More general Riemannian analogies have been introduced by F. Otto (see Otto (2001)).

For completeness reasons, based on the above characterizations, let us present, without proof, some important results related with Wasserstein distances, which will add some extra hints about the Riemannian structure of Wasserstein spaces. Its proof can be found in L. Ambrosio et al. (2008).

Proposition 4.4.4. *Consider that $(\mu_t)_{t\in[0,1]}, (\nu_t)_{t\in[0,1]} \subset \mathbb{W}_2(\mathbb{R}^d)$ are defined as Proposition 1.2.2 and $\nu \in \mathbb{W}_2(\mathbb{R}^d)$. Then:*

(i) v_t can be recovered by infinitesimal displacements, that is, if $\boldsymbol{T}_t^s$ is the optimal transport map from μ_t to μ_s, then it holds

$$v_t = \lim_{s\to t} \frac{\boldsymbol{T}_t^s - \mathsf{Id}}{s-t}, \qquad \textit{for a.e. } t \in [0,1].$$

(ii) we have "displacement tangency", that is,

$$\lim_{h\to 0} \frac{\mathcal{W}_2(\mu_{t+h}, (\mathsf{Id} + hv_t)_{\#}\mu_t)}{h} = 0$$

(iii) the derivative of squere Wasserstein distance has the form

$$\frac{d}{dt}\mathcal{W}_2^2(\mu_t, \nu) = -2\int_{\mathbb{R}^d} \langle u_t, T_t - \mathsf{Id}\rangle d\mu_t,$$

where $\boldsymbol{T}_t$ is the optimal transport map between μ_t and ν.

CHAPTER 5

Gradient flows in metric spaces

"Problems worthy of attack prove their worth by fighting back"

— Paul Erdös,
Hungarian mathematician (1913-1996)

In this chapter we present the last scene our exploration, which was essentially the goal of this thesis: gradient flows in metric spaces. Having as starting point the metric variational characterizations as well as the approximation ideas of Minimizing movements scheme, which were described in Chapter 1 and employing the theory of curves in metric spaces, we will explore the gradient flows in metric space setting. Moreover, we will present four groups of assumptions, which will make us able to recast many important results, such as existence, uniqueness and contractivity. All of our results can be found in the *bible* of gradient flows of L. Ambrosio, N. Gigli and G. Savare (L. Ambrosio et al. (2008)) or in the very well-written lectures notes of L. Ambrogio and N. Gigli (Ambrosio L., Gigli N. (2013)) and the expository paper of F. Santambrogio (**?**), in detail. Let us mention the following disclaimer: during this chapter, for simplicity, for a given $u \in \mathcal{X}$, shall we denote as u_t, the dependence of u on variable t. This will not be true for geodesics and constant speed geodesics, where their variable dependence will still be denoted as $\gamma(t)$.

5.1 Metric characterizations-revisited

As it was presented in Section 1.4, living in Euclidean space, we were able to prove three important metric variational characterizations of gradient flows, that is Energy Dissipation Equality (EDE), Energy Dissipation Inequality (EDI) and Evolution Variational Inequality (EVI). All of these characterizations had provide us interesting features related with the nature of gradient flows in Euclidean space, where we just ask only λ-convexity and smoothness assumptions of functional ϕ. So, one may ask *can we recast all of these characterizations in metric setting, in order to extend the notion of gradient flows?*

At first glance, the answer might be negative, since they involved the differential structure of Euclidean space, which actually lies to its vector nature. Nevertheless, under the light of the discussion which was presented in Chapter 2, and based on quite weak and reasonable assumptions, we are able to recast differentiability notions in metric space setting, like metric derivative and metric slope. Thus, having the theory of curves in metric spaces in our toolbox, naively, it seems to be possible to explore all of these metric variational characterizations in *metric setting,* instead of the classical vector space setting which we have already discussed. So, under this considerations, and based to these variational characterizations, let us present three notions of gradient flows beyond vectorial setting.

Definition 5.1.1 (Gradient flow-Energy Dissipation Inequality sense)**.** Consider that $(\mathcal{X}, d)$ is a metric space, $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a functional and $\overline{u} \in \mathcal{X}$ such that $\phi(\overline{u}) < \infty$. Then, we will say that $u_t \in AC_{loc}((0,\infty); \mathcal{X})$ is a gradient flow, in Energy Dissipation Inequality (EDI) sense, starting at $\overline{u}$ with $u_0 = \overline{u}$ and

$$\phi(u_s) + \frac{1}{2}\int_0^s |u_r'|^2 dr + \frac{1}{2}\int_0^s |\nabla\phi(u_r)|^2 dr \leq \phi(\overline{u}), \quad \text{for every } s \geq 0 \tag{5.1}$$

$$\phi(u_s) + \frac{1}{2}\int_t^s |u_r'|^2 dr + \frac{1}{2}\int_s^t |\nabla\phi(u_r)|^2 dr \leq \phi(u_t), \quad \text{a.e. } t > 0, \text{ and for every } s \geq t. \tag{5.2}$$

Definition 5.1.2 (Gradient flow-Energy dissipation equality sense)**.** Consider that $(\mathcal{X}, d)$ is a metric space, $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a functional and $\overline{u} \in \mathcal{X}$ such that $\phi(\overline{u}) < \infty$. Then, we will say that $u_t \in AC_{loc}((0,\infty); \mathcal{X})$ is a gradient flow, in Energy Dissipation Equality (EDE) sense, at $\overline{u}$ with $u_0 = \overline{u}$ and

$$\phi(u_s) + \frac{1}{2}\int_t^s |u_r'|^2 dr + \frac{1}{2}\int_s^t |\nabla\phi(u_r)|^2 dr = \phi(u_t), \quad \text{for every } t \in [0, s].$$

Definition 5.1.3 (Gradient flow-Evolution variational inequality sense)**.** Consider that $(\mathcal{X}, d)$ is a metric space, $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a functional and $\overline{u} \in \mathcal{X}$ such that $\phi(\overline{u}) < \infty$. Then, we will say that $u_t \in AC_{loc}((0,\infty); \mathcal{X})$ is a gradient flow, in Evolution Variational Inequality (EVI) sense, with respect λ, starting at $\overline{u}$, where $u_t \to \overline{u}$ as $t \to 0$ and

$$\phi(u_t) + \frac{1}{2}\frac{d}{dt} d^2(u_t, y) + \frac{\lambda}{2} d^2(u_t, v) \leq \phi(v), \quad \text{for every } v \in \mathcal{X} \text{ and a.e. } t > 0.$$

We have seen that all of these characterizations are equivalent in a linear environment for λ-convex functionals. An interesting question arising from this observation is: is that also true in metric space setting? Unfortunately, the answer is negative. In particular, we have the following implication: Energy Dissipation equality implies Energy dissipation Inequality, and Evolution Variational inequality implies Energy Dissipation equality. Although, none of above converse implications is true. We will see a specific pathological case in Example 5.2.9). Before of exploring these converse implications, let us present the latter implication of these notions of gradient flows in the following proposition.

Proposition 5.1.4 (From EVI to EDE)**.** *Consider that $(\mathcal{X}, d)$ is a metric space and $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous functional. Given a point $\overline{u} \in \mathcal{X}$ and $\lambda \in \mathbb{R}$, we assume that $u_t \in AC_{loc}((0,\infty); \mathcal{X})$ is a gradient flow in Evolutional Variational Inequality (EVI) sense starting at point $\overline{u}$. Then it holds*

$$\phi(u_s) + \frac{1}{2}\int_t^s |u_r'| dr + \frac{1}{2}\int_s^t |\nabla\phi(u_r)|^2 dr = \phi(u_t), \quad \textit{for } 0 \leq t \leq s.$$

Proof. For simplicity, let us assume that $u_t \in AC_{loc}((0,\infty); \mathcal{X})$ is locally Lipschitz. Given this, the statement of the proposition will be proved if we prove that the map $t \mapsto \phi(u_t)$ is also locally Lipschitz and it holds that

$$-\frac{d}{dt}\phi(u_t) = \frac{1}{2}|u_t'|^2 + \frac{1}{2}|\nabla\phi(u_t)|^2, \quad \text{a.e. for } t > 0. \tag{5.3}$$

Note, that the assumption that u_t is locally Lipschitz is not implausible. Let us sketch the argument of the proof of this statement. At first, one can observe that the map $t \mapsto u_{t+h}$ is a gradient flow in Evolutional Variational Inequality sense, starting from u_h for some $h > 0$. Now, thanks to Corollary 4.3.3. of L. Ambrosio et al. (2008), the distance between these curves, satisfying Evolutional Variational Inequality, is contractive up to an exponential factor. Thus we have

$$d(u_s, u_{s+h}) \leq e^{\lambda(s-t)} d(u_t, u_{t+h}), \text{ for every } s > t \text{ and for some } \lambda \in \mathbb{R}. \tag{5.4}$$

Now, calling $B \subset (0, \infty)$, the set where the metric derivative of $|u'_t|$ exists, dividing (5.4) by h, letting h travel to zero, we obtain

$$|u'_s| \leq |u'_t| e^{\lambda(s-t)}, \qquad \text{for every } s, t \in B \text{ with } s > t,$$

from which we infer that u_t is Lipschitz in $(0, \infty)$.

Back to the proof of the statement of the proposition, one can observe that thanks to the triangle inequality, we obtain

$$\frac{1}{2}\frac{d}{dt} d^2(u_t, v) \geq -|u'_t| d(u_t, v), \qquad \text{for every } v \in \mathcal{X} \text{ and a.e. } t > 0. \tag{5.5}$$

Plugging the bound given in (5.5) in the Evolution Variational Inequality we obtain

$$-|u'_t| d(u_t, v) + \frac{\lambda}{2} d^2(u_t, v) + \phi(u_t) \leq \phi(v), \qquad \text{for every } v \in \mathcal{X} \text{ and a.e. } t > 0,$$

which implies

$$|\nabla \phi(u_t)| = \limsup_{v \to u_t} \frac{\big(\phi(u_t) - \phi(v)\big)^+}{d(u_t, v)} \leq |u'_t|, \qquad \text{a.e. } t > 0. \tag{5.6}$$

We choose an interval $[a, b] \subset (0, +\infty)$ and consider that L is the Lipschitz constant of u_t in $[a, b]$. Then for any $v \in \mathcal{X}$, it holds that

$$\frac{d}{dt} d^2(u_t, v) \geq -|u'_t| d(u_t, v) \geq -L d(u_t, v), \qquad \text{a.e. } t \in [a, b]. \tag{5.7}$$

Plugging again the bound given by (5.7) in Evolution Variational Inequality, we obtain

$$-L d(u_t, v) + \frac{\lambda}{2} d^2(u_t, v) + \phi(u_t) \leq \phi(v), \qquad \text{a.e. } t \in [a, b]. \tag{5.8}$$

Thanks to the lower semicontinuity condition on ϕ, we have that the inequality (5.8) is also true *for every* $t \in [a, b]$. Taking $y = u_s$, for some $s \in [a, b]$, and exchanging the roles of u_t and u_s, we deduce that

$$\begin{aligned} |\phi(u_t) - \phi(u_s)| &\leq L d(u_t, u_s) - \frac{\lambda}{2} d^2(u_t, u_s) \\ &\leq L|t - s| \left(L + \frac{|\lambda|}{2} L|t - s| \right), \qquad \text{for every} t, s \in [a, b], \end{aligned}$$

thus the map $t \mapsto \phi(u_t)$ is locally Lipschitz. Then, we obseve that

$$\begin{aligned} -\frac{d}{dt}\phi(u_t) &= \lim_{h \to 0} \frac{\phi(u_t) - \phi(u_{t+h})}{h} \\ &= \lim_{h \to 0} \frac{\phi(u_t) - \phi(u_{t+h})}{d(u_{t+h}, u_t)} \cdot \frac{d(u_{t+h}, u_t)}{h} \\ &\leq |\nabla \phi(u_t)| \cdot |u'_t| \\ &\leq \frac{1}{2}|\nabla \phi(u_t)|^2 + \frac{1}{2}|u'_t|^2, \qquad \text{a.e. } t > 0. \end{aligned}$$

Therefore, it suffice now to prove the converse inequality. To do this, we integrate the Evolution Variational Inequality in $[t, t+h]$, and we obtain

$$\frac{d^2(u_{t+h}, v) - d^2(u_t, v)}{2} + \int_t^{t+h} \phi(u_s)ds + \frac{\lambda}{2}\int_t^{t+h} d(u_s, v)ds \leq h\phi(v), \quad \text{for every } v \in \mathcal{X} \text{ and a.e. } t > 0.$$

In the same fashion as we before, letting $v = u_t$, we get

$$\frac{d^2(u_{t+h}, u_t)}{2} \leq \int_t^{t+h} \phi(u_t) - \phi(u_s)ds + \frac{|\lambda|}{6}L^2h^3 = h\int_0^1 \phi(u_t) - \phi(u_{t+hr})dr + \frac{|\lambda|}{6}L^2h^3. \tag{5.9}$$

Consider now $A \subset (0, +\infty)$, the set where the map $t \mapsto \phi(u_t)$ is differentiable and the metric derivative $|u_t'|$ exists. Choosing $t \in A \cap (a, b)$, and dividing the inequality (5.9) by h^2, taking the limit as h goes to zero and using Lebesgue's dominated convergence theorem, we obtain

$$\frac{1}{2}|u_t'|^2 \leq \lim_{h\to 0}\int_0^1 \frac{\phi(u_t) - \phi(u_{t+hr})}{h}dr = -\frac{d}{dt}\phi(u_t)\int_0^1 rdr = \frac{1}{2}\frac{d}{dt}\phi(u_t). \tag{5.10}$$

Conbining now (5.6) and (5.10), we obtain

$$-\frac{d}{dt}\phi(u_t) \geq \frac{1}{2}|u_t'|^2 + \frac{1}{2}|\nabla\phi(u_t)|^2, \quad \text{a.e. } t > 0,$$

and thus we obtain the equality, which concludes our proof. □

5.2 Existence in the metric setting

Having all of these three notions of gradient flows in metric spaces in our mind, perhaps the first and most important questions arising, in this setting, is related with their existence. How we can prove the existence of them? In order to tackle such a concern we will employ the *Minimizing Movements* scheme, which was discussed in Remark 1.3.2 in Euclidean setting. Heuristically, the idea of the programme that will follow is quite simple: *discretize in time and then find the suitable assumptions to pass to the limit.* However, as we will see in the following, there are some technical difficulties. Let us make a step further beyond Euclidean setting, in order to dis-mystify these technicalities. For that reason let us consider that $\mathcal{H}$ is separable Hilbert space, and $\phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a convex and lower semicontinuous functional. In the same fashion as Euclidean setting, we pick a $\overline{u} \in \overline{D(\phi)} = \overline{\{u \in \mathcal{H} : \phi(u) < \infty\}}$, we fix $\tau > 0$ and define the sequence $(u_n^\tau)_{n\in\mathbb{N}}$, recursively by setting $u_n^\tau := \overline{u}$ and defining u_{n+1}^τ as a minimizer of

$$u \mapsto \phi(x) + \frac{|u - u_n^\tau|^2}{2\tau}.$$

Up to a direct method of Calculus of Variations argument, we can observe that the minimizer is unique, and thus the sequence $(u_n^\tau)_{n\in\mathbb{N}}$ is well-defined. The Euler-Lagrange equation of u_{n+1}^τ is

$$\frac{u_{n+1}^\tau - u_n^\tau}{\tau} \in -\partial\phi(u_{n+1}^\tau)$$

and is a typical time-discretization of the gradient flow system of problem.

Back to metric setting, and imitating the above discussion, it is natural to introduce the rescaled curve $(u_t^\tau)_{t\in(0,+\infty)}$ by setting

$$u_t^\tau := u_{[t/\tau]}^\tau.$$

Having this time-discretization procedure in our mind, a natural question about this discretization scheme is whether the curves $(u_t^\tau)_{t\in(0,+\infty)}$ convergence, in some sense and under certain assumptions, to a limit curve which solves the system of gradient flow, as $\tau \downarrow 0$.

The generality of the above scheme could be very important in the following. Precisely, this minimization problem, due to its nature, can be posed in a purely metric setting for a general functional $\phi : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$. To see this, it is sufficient to pick a $\overline{u} \in \overline{D(\phi)}$, consider $\tau > 0$, define $u_0^\tau = \overline{u}$ and recursively solve

$$u_{n+1}^\tau \in \operatorname{argmin}\left\{\phi(u) + \frac{d^2(u_n^\tau, u)}{2\tau}\right\}.$$

This observation motivates the definition of *discrete solution,* which is a key ingredient of Minimizing Movements scheme.

Definition 5.2.1 (Discrete solution)**.** Consider that $(\mathcal{X}, d)$ is a complete and separable metric space and $\phi : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ be a lower semicontinuous functional. Consider also $\overline{u} \in \overline{D(\phi)}$ and $\tau > 0$. A discrete solution is a map $u_t^\tau : [0, +\infty) \to \mathcal{X}$ defined by $u_t^\tau := u_{[t/\tau]}^\tau$, where $u_0^\tau := \overline{u}$ and

$$u_{n+1}^\tau \in \operatorname{argmin}\left\{\phi(u) + \frac{d^2(u_n^\tau, u)}{2\tau}\right\} \tag{5.11}$$

Clearly, in the metric space setting, it is important to identify of suitable assumptions that ensure that the minimization problem (5.11) admits at least a solution, so these discrete solutions exist. Now, we will see under what conditions on functional ϕ or the space $(\mathcal{X}, d)$ is it possible to prove the existence of solutions in that Energy Dissipation Inequality, Energy Dissipation Equality or Evolution Variational Inequality sense.

Essentially, there are two classes of assumptions which we have to keep in mind. The first class consists of assumptions that can ensure the existence of discrete solutions under all of above sense. The second class consists of assumptions that can guarantee that we can pass to the limit. We shall now present the first class.

Assumptions 5.2.2. Consider that $(\mathcal{X}, d)$ is complete and separable metric space and $\phi : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ a lower semicontinuous and bounded from below functional. We assume that there exists a $\overline{\tau} > 0$ such that for every $0 < \tau < \overline{\tau}$ and $\overline{u} \in \overline{D(\phi)}$ there exists at least a minimum of the functional

$$x \mapsto \phi(u) + \frac{d^2(u, \overline{u})}{2\tau} \tag{5.12}$$

Thanks to Assumptions 5.2.2 we know that discrete solutions exists for every starting point $\overline{u} \in \overline{D(\phi)}$ for sufficient small τ. The big problem is to show that the discrete solutions satisfy the discretized version of Energy Dissipation Inequality in order to pass to the limit. The crucial fact which enable us to do this comes from E. de Giorgi legacy and its described in the following proposition.

Proposition 5.2.3 (E. De Giorgi)**.** *Consider that $(\mathcal{X}, d)$ is a metric space and a functional $\phi : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ which satisfies Assumptions 5.2.2. We fix a $\overline{u} \in \mathcal{X}$ and for any $0 < \tau < \overline{\tau}$, we choose u_τ among the minimizers of 5.12. Then the map*

$$\tau \mapsto \phi(u_\tau) + \frac{d^2(u, u_\tau)}{2\tau}$$

is locally Lipschitz in $(0, \overline{\tau})$ and it holds

$$\frac{d}{dt}\left(\phi(u_\tau) + \frac{d^2(u, u_\tau)}{2\tau}\right) = -\frac{d^2(u, u_\tau)}{2\tau^2}, \qquad a.e.\ \tau \in (0, \overline{\tau}).$$

Proof. By a straightforward observation, we can see that for $\tau_0, \tau_1 \in (0, \tau)$ with $\tau_0 < \tau_1$, it holds

$$\phi(u_{\tau_0}) + \frac{d^2(u_{\tau_0}, u)}{2\tau_0} \leq \phi(u_{\tau_1}) + \frac{d^2(u_{\tau_1}, u)}{2\tau_1}. \tag{5.13}$$

Then, we can deduce that

$$\phi(u_{\tau_0}) + \frac{d^2(u_{\tau_0}, u)}{2\tau_0} - \phi(u_{\tau_1}) + \frac{d^2(u_{\tau_1}, u)}{2\tau_1} \leq \left(\frac{1}{2\tau_0} - \frac{1}{2\tau_1}\right)\phi^2(u_{\tau_1}, u) = \frac{\tau_1 - \tau_0}{2\tau_0\tau_1} d^2(u_{\tau_1}, u).$$

In the same fashion working symmetrically, we obtain

$$\phi(u_{\tau_0}) + \frac{d^2(u_{\tau_0}, u)}{2\tau_0} - \phi(u_{\tau_1}) + \frac{d^2(u_{\tau_1}, u)}{2\tau_1} \geq \frac{\tau_1 - \tau_0}{2\tau_0\tau_1} d^2(u_{\tau_0}, u).$$

Based on these calculations, the map $u \mapsto \phi(u_\tau) + \frac{d^2(u, u_\tau)}{2\tau}$ is locally Lipschitz and we can also see that the desired equation is true, which concludes our proof. $\square$

So, since now, we have a first important existence result in order to study the gradient flow problem in metric space setting. To make a further step in this direction, let us present a useful bound for the metric slope, in the following lemma.

Lemma 5.2.4 (Metric slope bound)**.** *Under the assumptions of Proposition 5.2.3, the map $\tau \to d(u_\tau, \overline{u})$ is non-decreasing and it holds*

$$|\nabla\phi(x_\tau)| \leq \frac{d(x_\tau, \overline{x})}{\tau}. \tag{5.14}$$

Proof. Let $\tau_0, \tau_1 \in (0, \overline{\tau})$ with $\tau_0 < \tau_1$. Thanks to the minimality of u_{τ_0} and u_{τ_1}, we obtain

$$\phi(u_{\tau_0}) + \frac{d^2(u_{\tau_0}, \overline{u})}{2\tau_0} \leq \phi(u_{\tau_1}) + \frac{d^2(u_{\tau_1}, \overline{u})}{2\tau_0} \tag{5.15}$$

$$\phi(u_{\tau_1}) + \frac{d^2(u_{\tau_1}, \overline{u})}{2\tau_1} \leq \phi(u_{\tau_0}) + \frac{d^2(u_{\tau_0}, \overline{u})}{2\tau_1} \tag{5.16}$$

Since $\tau_0^{-1} + \tau_1^{-1} \geq 0$, adding (5.15) and (5.16), we can see that $d(u_{\tau_0}, \overline{u}) \leq d(u_{\tau_1}, \overline{u})$. By a rearrangement of the quantities in both sides, we can obtain that

$$\phi(u_{\tau_1}) + \frac{d^2(u_{\tau_0}, \overline{u})}{2\tau_1} \leq \phi(u_{\tau_1}) + \frac{d^2(u_{\tau_1}, \overline{u})}{2\tau_1} \leq \phi(u_{\tau_0}) + \frac{d^2(u_{\tau_0}, \overline{u})}{2\tau_1},$$

which proves that the map $\tau \mapsto \phi(u_\tau)$ is non decreasing.

The bound of metric slope comes from the observation that it holds

$$\phi(u_\tau) + \frac{d^2(u_\tau, \overline{u})}{2\tau} \leq \phi(y) + \frac{d^2(y, \overline{u})}{2\tau}, \quad \text{for every } y \in \mathcal{X}.$$

Based on this observation, we have that

$$\begin{aligned} \frac{\phi(u_\tau) - \phi(y)}{d(u_\tau, y)} &\leq \frac{d^2(y, \overline{u}) - d^2(x_\tau, \overline{u})}{2\tau d(u_\tau, y)} = \frac{\big(d(y, \overline{u}) - d(u_\tau, \overline{u})\big)\big(d(u_\tau, \overline{u}) + d(y, \overline{u})\big)}{2\tau d(u_\tau, y)} \\ &\leq \frac{d(u_\tau, \overline{u}) + d(y, \overline{u})}{2\tau} \end{aligned}$$

and thus,

$$\limsup_{y\to u_\tau}\frac{\phi(u_\tau)-\phi(y)}{d(u_\tau,y)}\le\limsup_{y\to u_\tau}\frac{d(u_\tau,\overline{u})+d(y,\overline{u})}{2\tau}$$

which implies

$$|\nabla\phi(u_\tau)|\le\frac{d(u_\tau,\overline{u})}{\tau}.$$

This fact makes our proof complete. □

In the same fashion as for the Euclidean setting and having the existence of discrete solutions in our toolbox, we can define a suitable notion of interpolation in the metric setting, imitating the usual constant interpolation. More precisely, we introduce the following variational interpolation in Minimizing Movements scheme, as opposed to the classical piecewise affine (or even constant) interpolations which are commonly used in many other problems with similar numerical nature. Moreover, in the same spirit of the new notion of interpolation, we define the "discretized" versions of speed and slope with respect to a time step parameter τ. Both of them are presented in what follows.

Definition 5.2.5 (Variational interpolation)**.** Consider that $(\mathcal{X},d)$ is metric space and $\phi:\mathcal{X}\to\mathbb{R}\cup\{+\infty\}$ is a functional satisfying Assumptions 5.2.2, $\overline{u}\in\overline{D(\phi)}$ and $0<\tau<\overline{\tau}$. We define the map $u^\tau(t):[0,+\infty)\to\mathcal{X}$ satisfying

(i) $u^\tau_0:=\overline{u}$

(ii) $u^\tau_{(n+1)\tau}$ is chosen among the minimizers of (5.12) with $\overline{u}$ replaced by $u^{n\tau}_\tau$

(iii) $u^\tau(t)$ with $t\in(n\tau,(n+1)\tau)$ is chosen among the minimizers of 5.12 with $\overline{u}$ and τ are replaced by $u^\tau_{n\tau}$ and $t-n\tau$ respectively.

For such a map $u^\tau(t)$, we define the discrete speed $\mathsf{dsp}^\tau:[0,+\infty)\to[0,+\infty)$ and the discrete slope $\mathsf{dsl}^\tau:[0,+\infty)\to[0,+\infty)$ as

$$\mathsf{dsp}^\tau_t:=\frac{d(u^\tau_{n\tau},u^\tau_{(n+1)\tau})}{\tau}\quad\text{and}\quad\mathsf{dsl}^\tau_t:=\frac{d(u^\tau_t,u^\tau_{n\tau})}{t-n\tau}\quad\text{for } t\in(n\tau,(n+1)\tau).$$

Let us note an important byproduct of the above definition. Based on the definition of discrete slope, we can understand that it stands for a completely different quantity that the metric slope. Nevertheless, after a some calculations and having the metric slope bound by Lemma 5.2.4 in our toolbox, we prove that $|\nabla\phi(u^\tau_t)|\le\mathsf{dsl}^\tau_t$. Thus, taking the limiting process, dsl^τ produces the slope term of gradient flow in Energy Dissipation Inequality sense, which justifies the term "discrete slope". We state the latter observation in the following corollary.

Corollary 5.2.6. *Consider that $(\mathcal{X},d)$ is a metric space and $\phi:X\to\mathbb{R}\cup\{+\infty\}$ satisfying the Assumptions 5.2.2, $\overline{u}\in\overline{D(\phi)}$, $0<\tau<\overline{\tau}$, and u^τ_t defined via variational interpolation. Then it holds*

$$\phi(u^\tau_s)+\frac{1}{2}\int_t^s|\mathsf{dsp}^\tau_r|^2dr+\frac{1}{2}\int_t^s|\mathsf{dsl}^\tau_r|^2dr=\phi(u^\tau_t),\tag{5.17}$$

for every $t=n\tau$ and $s=m\tau$, with $n<m\in\mathbb{N}$.

Thus, under quite general assumptions, at the level of discrete solutions, it is possible to obtain a discrete form of the Energy Dissipation Inequality. Although, this procedure sheds the light on a fundamental question on our time discretization scheme: can we pass to the limit? The answer is positive, but since there is no free lunch, we have to recall a class of several compactness and regularity-type assumptions of the functional ϕ. The latter class of assumptions is presented in the following.

Assumptions 5.2.7. We assume that the functional $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ satisfies:

(i) ϕ is bounded from below and its sublevel sets are bounded and compact i.e. the set

$$\{u \in \mathcal{X} : \phi(u) \leq c\} \cap \overline{B_r(u)}$$

is compact for any $c \in \mathbb{R}$ and $r > 0$.

(ii) the metric slope $|\nabla\phi| : D(\phi) \to [0, +\infty]$ is lower semicontinuous

(iii) ϕ satisfies the following continuity-compactness property

$$\text{if} \quad u_n \to u \quad \text{with} \quad \sup_{n\in\mathbb{N}} \big\{|\nabla\phi(u_n)|, \phi(u_n)\big\} < \infty \quad \text{then} \quad \phi(u_n) \longrightarrow \phi(u).$$

Based on Assumptions 5.2.7, we can obtain the Energy Dissipation Inequality characterization of gradient flow, as it witness the following proposition.

Proposition 5.2.8. *Consider that $(\mathcal{X}, d)$ is a complete and separable metric space and $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ satisfies Assumptions 5.2.7 and 5.2.7. Let also $\overline{u} \in \overline{D(\phi)}$ and $0 < \tau < \overline{\tau}$ and define the discrete solutions via variational interpolation. Then:*

(i) the set of curves $\{u_t^\tau : \tau \geq 0\}$ is relatively compact in the set of absolutely continuous curves in $\mathcal{X}$ with respect to topology of local uniform convergence.

(ii) any limit curve $(u_t)_{t\in[0,+\infty)}$ is a gradient flow in the Energy Dissipation Inequality sense.

Proof. (i) Thanks to Corollary 5.2.6 we have that

$$d^2(u_t^T, \overline{u}) \leq \left(\int_0^T |\mathsf{dsp}_r^\tau| dr\right)^2 \leq T \int_0^T |\mathsf{dsp}_r^\tau|^2 dr \leq 2T(\phi(\overline{u}) - \inf_{u\in\mathcal{X}} \phi(u)), \tag{5.18}$$

for every $t \leq T$, where $T = n\tau$. Therefore, for any $T > 0$, the set $\{u_t^\tau : t \leq T\}$ is uniformly bounded in τ. Since this set is contained in $\{u \in \mathcal{X} : \phi(u) \leq \phi(\overline{u})\}$, it is relatively compact. The fact that it is also relatively compact with respect to the topology of uniform convergence comes from a standard application of the Arzelá-Ascoli theorem, based on the inequality

$$d^2(u_t^\tau, u_s^\tau) \leq \left(\int_s^t \mathsf{dsp}_r^\tau dr\right)^2 \leq 2(s-t)(\phi(\overline{u}) - \inf_{u\in\mathcal{X}} \phi(u)), \tag{5.19}$$

for every $t = n\tau$ and $s = m$, with $n < m \in \mathbb{N}$. Now, let $(\tau_n)_{n\in\mathbb{N}}$ with $\tau_n \downarrow 0$ such that $(u_t^{\tau_n})_{n\in\mathbb{N}}$ converges locally uniformly to a limit curve u_t. Then, by standard arguments and based on inequality (5.19), one can check that the map $t \mapsto u_t$ is locally absolutely continuous and satisfies

$$\int_t^s |u_r'|^2 dr \leq \liminf_{n\to\infty} \int_t^s |\mathsf{dsp}_r^{\tau_n}|^2 dr, \qquad \text{for } 0 \leq t \leq s. \tag{5.20}$$

Then by the lower semicontinuity of $|\nabla\phi|$ and the slope bound given by (5.2.4), we obtain

$$|\nabla\phi(u_t)| \leq \liminf_{n\to\infty} |\nabla\phi(u_t^{\tau_n})| \leq \liminf_{n\to\infty} \mathsf{dsl}_t^{\tau_n} dr, \qquad \text{for every } t \in [0, +\infty).$$

Thus, employing Fatou's lemma, we have that

$$\int_t^s |\nabla\phi(u_r)|^2 dr \leq \int_t^s \liminf_{n\to\infty} |\nabla\phi(u_r^\tau)|^2 dr \leq \liminf_{n\to\infty} \int_t^s |\mathsf{dsl}_r^{\tau_n}|^2 dr \leq 2T\big(\phi(\overline{u}) - \inf_{u\in\mathcal{X}} \phi(u)\big). \tag{5.21}$$

If we write (5.17) for $t = 0$ and pass to the limit, we obtain the Energy Dissipation Inequality characterization of gradient flow. In addition, if we define the function

$$f(t) := \liminf_{n\to\infty} |\nabla\phi(u_t^{\tau_n})|$$

we can see that $\|f\|_{L^2([0,\infty))} < \infty$. Therefore, the set $A := \{t \in [0,\infty) : f(t) < \infty\}$ has full Lebesgue measure. Based on this fact, for every $t \in A$, we can find a subsequence $(\tau_{n_k})_{k\in\mathbb{N}}$ such that $\tau_{n_k} \downarrow 0$ such that $\sup_{k\in\mathbb{N}} |\nabla\phi(u_t^{\tau_{n_k}})| < \infty$. Using the continuity-compactness assumption, which was given in (iii) of Assumptions 5.2.7, we obtain that $\phi(u_t^{\tau_{n_k}}) \to \phi(u_t)$. Moreover, thanks to lower semicontinuity of ϕ, we have that $\phi(u_s) \leq \liminf_{k\to\infty} \phi(u_s^{\tau_{n_k}})$, for any $s \leq t$. So using (5.21) and (5.20), and taking limit in (5.17) as $\tau_{n_k} \downarrow 0$, we obtain

$$\phi(u_s) + \frac{1}{2}\int_t^s |u_r'|^2 dr + \frac{1}{2}|\nabla\phi(u_r)|^2 dr \leq \phi(u_t), \text{ for every } t \in A \text{ and for every } s \geq t,$$

which concludes our proof. □

In general, there are situations where we have no hope obtain equality in Energy Dissipation Inequality. The core component of such a pathology is purely technical and stands to the fact that we don't know whether the map $t \mapsto \phi(u_t)$ is an absolutely continuous map. In the following example, we construct such a case based on the singularities of Cantor function.

Example 5.2.9 (No hope for equality)**.** Let $\mathcal{X} = [0,1]$ equipped with Euclidean metric, $\mathcal{C} \subset \mathcal{X}$ the Cantor set with null Lebesgue measure and a continuous and integrable function $f : [0,1] \to [0,+\infty]$ such that $f(x) = +\infty$ for any $x \in \mathcal{C}$ and $f(x)$ be smooth for any $x \in \mathcal{X}\backslash\mathcal{C}$. Consider also $g : [0,1] \to [0,1]$ be the Cantor function, i.e. satisfying $g(0) = 0$, $g(1) = 1$ and be constant in each of connected components of $\mathcal{X}\backslash\mathcal{C}$. Then we define the functionals $\phi, \tilde{\phi} : [0,1] \to \mathbb{R}$ by

$$\phi(u) := -g(u) - \int_0^u f(s)ds \quad \text{and} \quad \tilde{\phi}(u) := -\int_0^u f(s)ds.$$

It is now hard to verify that both ϕ and $\tilde{\phi}$ satisfy both Assumptions 5.2.2 and 5.2.7, since f is continuous which implies that $\nabla\phi$ and $\nabla\tilde{\phi}$ are continuous If we construct a gradient flow starting from $u(0) = 0$, we can check that Minimizing Movements scheme converges in both cases to two absolutely continuous curves u_t and $\tilde{u}_t$ respectively, and satisfying

$$u_t' = -|\nabla\phi(u_t)| \quad \text{and} \quad \tilde{u}_t' = -|\nabla\tilde{\phi}(u_t)|, \quad \text{a.e. for } t \in [0,1].$$

Then, we can see that $|\nabla\phi(u)| = |\nabla\tilde{\phi}(u)|$ for every $u \in [0,1]$, and thus, since f is smooth in $[0,1]\backslash\mathcal{C}$, both of the above Cauchy problems admit a unique solution. So if u_t and $\tilde{u}_t$ are the gradient flows, they *must* coincide. In particular, the effect of g cannot be seen via the gradient flow structure. Based on this observation, one can verify that there is an Energy Dissipation Inequality characterization for both gradient flows, but there is an Energy Dissipation Equality characterization only for the functional ϕ.

5.3 Geodesically convex functionals

As a next step, let us proceed to a further exploration of gradient flow notions employing a geometrical flavour in our study. More precisely, we will assume that the space $(\mathcal{X}, d)$ is complete and separable metric space and also *geodesic metric space,* with several compactness-type properties. Under this consideration, we will assume that the functional ϕ stands for λ-geodesically functional, recalling the related discussion of Chapter 2. To make this discussion a bit more rigorous, let us present all the necessary assumptions more formally.

Assumptions 5.3.1. Let $(\mathcal{X}, d)$ be a complete and separable geodesic metric space and $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous, λ-geodesically convex functional for some $\lambda \in \mathbb{R}$. We assume that the sublevel sets of ϕ are bounded and compact, i.e. the set $\{u \in \mathcal{X} : \phi(u) \leq c\} \cap \overline{B_r(u)}$ is compact for any $c \in \mathbb{R}, r > 0$ and $u \in \mathcal{X}$.

Using Assumptions 5.3.1, we are able to study gradient flow problem with a different and quite important perspective. Precisely, as we will see we will be able to obtain gradient flow in Energy Dissipation Equality sense, which as we have see in Example 5.2.9, was a little bit problematic.

Let us present some useful results.

Lemma 5.3.2 (Metric slope representation)**.** *Consider that $(\mathcal{X}, d)$ is a complete and separable geodesic metric space and $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ a λ-geodesically convex functional. Then it holds that*

$$|\nabla\phi(u)| = \sup_{v \neq u} \left(\frac{\phi(u) - \phi(v)}{d(u, v)} + \frac{\lambda}{2} d(u, v) \right)^+, \quad \textit{for every } u \in \mathcal{X}.$$

Proof. One can observe that, by the definition of metric slope, it holds

$$|\nabla\phi(u)| = \limsup_{v \to u} \left(\frac{\phi(u) - \phi(v)}{d(u, v)} + \frac{\lambda}{2} d(u, v) \right)^+ \leq \sup_{v \neq u} \left(\frac{\phi(u) - \phi(v)}{d(u, v)} + \frac{\lambda}{2} d(u, v) \right)^+.$$

To obtain the converse inequality, we fix $v \neq u$, and a constant speed geodesic $\gamma : [0, 1] \to \mathcal{X}$ as in definition of λ-geodesic convexity (see Definition 2.5.5). Then we have that

$$\begin{aligned}
|\nabla\phi(u)| &\geq \limsup_{t \downarrow 0} \left(\frac{\phi(u) - \phi(v)}{d(u, \gamma(t))} \right)^+ \\
&= \left(\limsup_{t \downarrow 0} \frac{\phi(u) - \phi(v)}{d(u, \gamma(t))} \right)^+ \\
&\geq \left(\limsup_{t \downarrow 0} \left(\frac{\phi(u) - \phi(v)}{d(u, \gamma(t))} + \frac{\lambda}{2}(1 - t) d(u, v) \right) \right)^+ \\
&\geq \sup_{v \neq u} \left(\frac{\phi(u) - \phi(v)}{d(u, v)} + \frac{\lambda}{2} d(u, v) \right)^+,
\end{aligned}$$

so the equality is true. □

The usefulness of the metric slope representation stands to the fact that it allows us to prove and interesting implication about the three classes of assumptions, which we have already presented. In other words, based on metric slope representation, we are able to show that Assumptions 5.3.1 implies both of Assumptions 5.2.2 and 5.2.7, as the following proposition presents.

Proposition 5.3.3. *If Assumptions 5.3.1 are true, then Assumptions 5.2.2 and 5.2.7 are also true.*

Proof. From λ-geodesically convexity and lower semicontinuity of ϕ we can deduce that it has at most quadratic decay at infinity, i.e. there exists $a, b > 0$ and $\overline{u} \in \mathcal{X}$ such that

$$\phi(u) \geq -a - b d(u, \overline{u}) + \lambda^- d^2(u, \overline{u}), \quad \text{for every } u \in \mathcal{X}.$$

Therefore, again from lower semicontinuity and bounded compactness of sublevel sets of ϕ, we can conclude that 5.12 admits a solutions whenever $\tau < 1/\lambda^-$.

The lower semicontinuity of slope comes directly from Lemma (5.3.2) and again by lower semicontinuity of ϕ. Thus, in order to prove the implication we have to prove only that the continuity-compactness property holds too, i.e.

$$\text{if} \quad u_n \to u \quad \text{with} \quad \sup_{n\in\mathbb{N}} \big\{|\nabla\phi|(u_n), \phi(u_n)\big\} < \infty \quad \text{then} \quad \limsup_{n\to\infty} \phi(u_n) \le \phi(u).$$

This can be obtained using Lemma 5.3.2, and replacing u, y by u_n, u respectively, i.e

$$\phi(u) \ge \phi(u_n) - |\nabla\phi|(u_n)d(u,u_n) + \frac{\lambda}{2}d^2(u,u_n).$$

Taking the limit as $n \to \infty$, we obtain the desired result. □

Based on above interesting implication, we are able to prove the Energy Dissipation Inequality formulation of gradient flow. Although, in order to obtain Energy Dissipation Equality formulations, we need a kind of weak chain rule, which could predict pathological cases as Example 5.2.9. Such type of weak chain rule, is presented in the following proposition.

Proposition 5.3.4. *Consider that $(\mathcal{X}, d)$ is a complete separable and geodesic space and $\phi : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ id λ-geodesically convex lower semicontinuous functional. Then for every $u_t \in AC((0,+\infty);\mathcal{X})$ such that for every $t \in (0,+\infty)$ it holds that $\phi(u_t) < \infty$, then we have*

$$|\phi(u_s) - \phi(u_t))| \le \int_t^s |u_r'||\nabla\phi|\big(u_r\big)dr, \qquad \textit{for every } t < s. \tag{5.22}$$

Proof. Let us assume that the right hand side of (5.22) is finite for any $s, t \in [0,1]$, and, up to a reparametrization argument which can achieved based on the discussion on the relevant results of Chapter 2 (Section 2.2.). We can also assume that $|u_t'| = 1$ and moreover $(u_t)_{t\in[0,1]}$ is 1-Lipschitz Thus, the map $t \mapsto |\nabla\phi\big(u_t\big)|$ belongs to $L^1((0,1))$. If the map $t \mapsto \phi\big(u_t\big)$ is absolutely continuous, then it holds

$$\begin{aligned}
\limsup_{h\uparrow 0} \frac{\phi(u_{t+h}) - \phi(u_t)}{h} &\le \limsup_{h\uparrow 0} \frac{\phi\big(u_{t+h}\big) - \phi(u_t)^+}{|h|} \\
&\le \limsup_{h\uparrow 0} \frac{\phi(u_{t+h}) - \phi(u_t)^+}{d(u_t,u_{t+h})} \limsup_{h\uparrow 0} \frac{d(u_t,u_{t+h})}{|h|} \\
&\le |\nabla\phi(u_t)||u_t'|,
\end{aligned}$$

for every $t \in [0,1]$ and implies the inequality (5.22). So our job is to prove the absolute continuity of $t \mapsto \phi\big(u_t\big)$. To do this, we will following an approximation procedure. We define the functions $f, g : [0,1] \to \mathbb{R}$ by

$$f(t) := \phi\big(u_t\big) \quad \text{and} \quad g(t) := \sup_{s\ne t} \frac{\big(f(t) - f(s)\big)^+}{|s-t|}.$$

Let also D be the diameter of the compact set $\{u_t : t \in [0,1]\}$. Now, since u_t is 1-Lipschitz, thanks to Lemma 5.3.2 and the fact that for any $a, b \in \mathbb{R}$ it holds that $a^+ \le (a+b)^+ + b^-$, we obtain that

$$g(t) \le \sup_{s\ne t} \frac{\big(\phi(u_t) - \phi(u_s)\big)^+}{d(u_s,u_t)} \le |\nabla\phi(u_t)| + \frac{\lambda^-}{2}D.$$

Thus, our claim we will be proved, if we prove the following implication

$$\text{if} \quad g \in L^1((0,1)) \quad \text{then} \quad |f(s) - f(t)| \le \int_t^s g(r)dr, \text{ for every } t < s. \tag{5.23}$$

To do this, we fix $M > 0$ and define $f^M = \min\{f, M\}$. Then we fix $\epsilon > 0$, and consider $\rho_\epsilon : \mathbb{R} \to \mathbb{R}$ be a the convolution kernel, supported in $[-\epsilon, \epsilon]$. Under these considerations, we define $f_\epsilon^M, g_\epsilon^M : [\epsilon, 1-\epsilon] \to \mathbb{R}$ by

$$f_\epsilon^M(t) := f^M * \rho_\epsilon(t) \quad \text{and} \quad g_\epsilon^M(t) := \sup_{s \neq t} \frac{\big(f_\epsilon^M(t) - f_\epsilon^M(s)\big)^+}{|s-t|}.$$

Then one can observe that by construction it holds $g_\epsilon^M \geq (f_\epsilon^M)'$, and since f_ϵ^M is smooth, it holds

$$|f_\epsilon^M(s) - f_\epsilon^M(t)| \leq \int_t^s g_\epsilon^M(r)dr. \tag{5.24}$$

Now, using the trivial fact that for a given function $h : [0,1] \to \mathbb{R}$ it holds $(\int h dx)^+ \leq \int h^+ dx$, we obtain

$$\begin{aligned}
g_\epsilon^M &\leq \sup_{s\in[\epsilon,1-\epsilon]} \frac{\displaystyle\int_s^t \big(f^M(t-r) - f^M(s-r)\big)^+ \rho_\epsilon(r)dr}{|s-t|} \leq \sup_{s\in[\epsilon,1-\epsilon]} \frac{\displaystyle\int_s^t \big(f(t-r) - f(s-r)\big)^+ \rho_\epsilon(r)dr}{|s-t|} \\
&= \sup_{s\in[\epsilon,1-\epsilon]} \frac{\displaystyle\int_s^t \big(f(t-r) - f(s-r)\big)^+ \rho_\epsilon(r)dr}{|(s-r)-(t-r)|} \leq \int_s^t g(t-r)\rho_\epsilon(r)dr \\
&= g * \rho_\epsilon(t).
\end{aligned}$$

Therefore, the family of functions $(g_\epsilon^M)_{\epsilon>0}$ is dominated in $L^1((0,1))$. Using (5.24) and the above domination, we can infer that the family of functions $(f_\epsilon^M)_{\epsilon>0}$ converges to a functions $\tilde{f}^M$ on $[0,1]$ with respect to the topology of uniform convergence as $\epsilon \downarrow 0$. For this limit it holds that

$$|\tilde{f}^M(s) - \tilde{f}^M(t)| \leq \int_t^s g(r)dr. \tag{5.25}$$

Although, by construction we have that $f^M = \tilde{f}^M$ on a set $A \subset [0,1]$ such that $\mathscr{L}^1\big([0,1]\backslash A\big) = 0$. Our purpose is to prove that actually they coincide everywhere. We remember that f^M is lower semicontinuous and its limit, i.e. $\tilde{f}^M$, is continuous and hence $f^M \leq \tilde{f}^M$ in $[0,1]$.

For the sake of contradiction, let assume that there exists a $t_0 \in [0,1]$ such that $f^M(t_0) < c < C < \tilde{f}^M(t_0)$, for some $c, C \in \mathbb{R}$. Then, we can find $\delta > 0$ such that $\tilde{f}^M(t) > C$ for $t \in [t_0-\delta, t_0+\delta]$. Therefore $f^M(t) > C$ for $t \in [t_0-\delta, t_0+\delta] \cap A$. Under this perspective, we have

$$\int_0^1 g(t)dt \geq \int_{[t_0-\delta,t_0+\delta]\cap A} g(t)dt \geq \int [t_0-\delta, t_0+\delta] \cap A \frac{C-c}{t-t_0} = +\infty,$$

which consists a contradiction, thanks to (5.25). Hence, we just proved that

$$\text{if} \quad g \in L^1((0,1)) \quad \text{then} \quad |f^M(t) - f^M(s)| \leq \int_t^s g(r)dr, \quad \text{for every } t < s \in [0,1] \text{ and } M > 0.$$

Passing to the limit as $M \to \infty$, we obtain the implication (5.23), which makes up proof complete. □

Now, based on previous proposition, we are ready to pass from existence of gradient flow in Energy Dissipation Inequality sense, to the existence in Energy Dissipation Equality sense.

Proposition 5.3.5. *Consider that $(\mathcal{X}, d)$ is a complete, separable and geodesic space and $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ satisfying the Assumptions 5.3.1. Let also $\overline{u} \in \overline{D(\phi)}$. Then, any gradient flow in the Energy Dissipation Inequality sense is also a gradient flow in the Energy Dissipation Equality sense.*

Proof. Thanks to Proposition 5.2.8, we have that the limit curve u_t is absolutely continuous and satisfies

$$\phi(u_s) + \frac{1}{2}\int_0^s |u_r'|^2 dr + \frac{1}{2}\int_0^s |\nabla\phi(u_r)| dr \leq \phi(\overline{u}), \qquad \text{for every } s \geq 0.$$

Moreover, the maps $t \mapsto |u_t'|$ and $t \mapsto |\nabla\phi(u_t)|$ belongs to $L^2_{loc}((0, +\infty))$. Thanks to Proposition 5.3.4, we have

$$|\phi(\overline{u}) - \phi(u_s) \leq \int_0^s |u_r'||\nabla\phi(u_r)| dr \leq \frac{1}{2}\int_0^s |u_r'|^2 dr + \frac{1}{2}\int_0^s |\nabla\phi(u_r)|^2 dr, \qquad \text{for any } s \geq 0.$$

Thus, the map $t \mapsto \phi(u_t)$ is absolutely continuous and it holds that

$$\phi(u_s) + \frac{1}{2}\int_0^s |u_r'|^2 dr + \frac{1}{2}\int_0^s |\nabla\phi(u_r)|^2 dr = \phi(\overline{u}). \tag{5.26}$$

Therefore, if we write (5.26) for $s = t$ and subtracting from itself, the desired result follows, and our proof is completed. □

At this point, we have to mention that in general λ-geodesically convexity hypothesis on functional ϕ does not play any important role about for what concerns compactness of the sequence of discrete solutions. Nevertheless, this hypothesis ensures several regularity properties for the limit curve. Some of theses properties are presented in the following proposition, which can be found in L. Ambrosio et al. (2008).

Proposition 5.3.6. *Consider that $(\mathcal{X}, d)$ is a complete, separable and geodesic metric space and $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ satisfying Assumptions 5.3.1. Consider also that u_t is any limit curve of a sequence of discrete solutions. Then the following holds:*

(i) the limit

$$|u_t'| := \lim_{h \downarrow 0} \frac{d(u_{t+h}, u_t)}{h}$$

exists for every $t > 0$.

(ii) the equation

$$\frac{d}{dt}\phi(u_t) = -|\nabla\phi(u_t)|^2 = -|u_t'|^2 = -|u_t||\nabla\phi(u_t)|,$$

is satisfied for every $t > 0$.

(iii) the map $t \mapsto e^{-2\lambda^- t}\phi(u_t)$ is convex, the map $t \mapsto e^{\lambda t}|\nabla\phi(u_t)|$ is non increasing, right continuous and satisfies

$$\frac{t}{2}|\nabla\phi(u_t)|^2 \leq e^{2\lambda^- t}\left(\phi(u_0) - \inf_{y\in\mathcal{X}}\left\{\phi(y) - \frac{d^2(u, y)}{2t}\right\}\right)$$

and

$$t|\nabla\phi(u_t)|^2 \leq (1 + 2\lambda^+ t)e^{-2\lambda t}\left(\phi(u_0) - \inf_{y\in\mathcal{X}}\left\{\phi(y) - \frac{d^2(u, y)}{2t}\right\}\right)$$

(iv) if $\lambda > 0$, then ϕ admits a unique minimum u^* and it holds

$$\frac{\lambda}{2}d^2(u_t, u^*) \leq \phi(u^*) - \phi(u^*) \leq e^{-2\lambda t}\big(\phi(u_0) - \phi(u^*)\big)$$

At this level of discussion, this observation is quite important, since it might be no hope for uniqueness of gradient flow in Energy Dissipation Equality sense without some additional assumptions. The following example present such a non-uniqueness case, quite less pathological than Cantorian case of no equality in Example 5.2.9.

Example 5.3.7 (No hope for uniqueness)**.** Let $\mathcal{X} = \mathbb{R}^2$ equipped with supremum norm, $\phi : \mathcal{X} \to \mathbb{R}$ defined as as $\phi(u_1, u_2) := u_1$ and initial datum $\overline{u} = (0, 0)$. Then, we can see that $|\nabla \phi| = 1$ and that any Lipschitz curve $t \mapsto u(t) = (u_1(t), u_2(t))$ satisfying

$$u_1(t) = -t, \quad \text{for every } t \geq 0 \quad \text{and} \quad |u_2'(t)| \leq 1 \quad \text{a.e. } t > 0,$$

satisfies also

$$\phi(u_t) = -t \quad \text{with} \quad |x'(t)| = 1.$$

Therefore, *any* such u_t satisfies Energy Dissipation Equality.

So, under the light of Example 5.3.7, in general uniqueness of the limit curve u_t obtain via Minimizing Movements scheme may collapse for a general λ-geodesically convex functional in Energy Dissipation Equality sense, which, as we have seen, is the weakest sense of a gradient flow. For that reason, in order to achieve properties like uniqueness and contractivity, we will introduce several assumptions between the functional ϕ and the distance d defined on $(\mathcal{X}, d)$ which *links* them in a stronger sense. These assumptions are imitating again convexity along to a geometrical flavor and they are described in the following.

Assumptions 5.3.8. Consider that $(\mathcal{X}, d)$ is a complete, separable and geodesic metric space, $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous functional and for any $u_0, u_1, v \in \mathcal{X}$, there exists a curve $(\gamma(t))_{t\in[0,1]}$ such that

$$\begin{aligned} \phi(\gamma(t)) &\leq (1-t)\phi(u_0) + t\phi(u_1) - \frac{\lambda}{2} t(1-t) d^2(u_0, u_1), && (5.27) \\ d^2(\gamma(t), v) &\leq (1-t) d^2(u_0, v) + t d^2(u_1, y) - t(1-t) d^2(u_0, u_1), && (5.28) \end{aligned}$$

for every $t \in [0, 1]$.

One can see that any compactness assumptions on sublevel sets of functional ϕ has been avoided. Although, in a Hilbert space setting, the inequality (5.28) is satisfied by geodesics. Therefore, λ-geodesically convex functionals are compatible with the metric ipso facto.

Adapting the same strategy as we have done previously, one can probe that Assumptions 5.3.8 implies 5.2.2. Nevertheless, in order to obtain the implication to Assumptions 5.2.7, we have to assume that the sublevels sets of ϕ satisfy some boundedness and compactness property. Then, thanks to Proposition 5.2.8, we can obtain gradient flow in Energy Dissipation Inequality sense. In addition, since under these assumptions the metric slope representation given by Lemma 5.3.2 holds, we can also obtain the gradient flow in Energy Dissipation Equality sense.

However, at this point, we will wipe the slate clean, and we will not follow the general theory which was developed previously. Under this consideration and based on Assumptions 5.3.8, we are able to prove much stronger results than those was developed in previous sections. Moreover, following the line of discrete solutions, we can prove the existence gradient flow in Evolution Variational Inequality sense, as it witnessed in the following proposition.

Proposition 5.3.9. *Consider that $(\mathcal{X}, d)$ is a complete, separable and geodesic metric space $(\mathcal{X}, d)$ and $\phi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a functional, satisfying Assumptions 5.3.8. Then the following hold*

(i) for every $\overline{u} \in \overline{D(\phi)}$ and $0 < \tau < 1/\lambda^-$ there exists a unique discrete solution u_t^τ

(ii) if $\overline{u} \in \overline{D(\phi)}$ is an initial datum and $(u_t^\tau)_{\tau>0}$ is a family of discrete solutions then $(u_t^\tau(t))_{\tau>0}$ converges to a unique limit curve u_t with respect to topology of uniform convergence as $\tau \downarrow 0$. Moreover, u_t is the unique solution of the system of differential inequalities

$$\frac{1}{2}\frac{d}{dt}d^2(\tilde{u}_t, y) + \frac{\lambda}{2}d^2(\tilde{u}_t, y) + \phi(\tilde{u}_t)) \leq \phi(y), \qquad a.e\ for\ t \geq 0\ and\ for\ every\ y \in \mathcal{X}, \tag{5.29}$$

among all absolutely continuous curves $\tilde{u}_t$ converging to $\overline{u}$ as $t \downarrow 0$, that is u_t is a gradient flow in Evolution Variational Inequality sense.

Proof. Without loss of generality, we assume that $\phi \geq 0$, $\lambda > 0$ and consider $\overline{u} \in \overline{D(\phi)}$.
(i) We will prove that the sequence of discrete solutions $(u_t^{\tau/2^n})_{n\mathbb{N}}$ converges to a limit curve u_t as $n \to \infty$ for any given $\tau > 0$.

To do this, we pick $u \in \mathcal{X}$, and we have to prove that there exists a unique minimizer for (5.12). Let $\mathcal{I} \geq 0$ be the infimum of (5.12) and $(u_n)_{n\in\mathbb{N}}$ be a minimizing sequence for (5.12). We fix $n, m \in \mathbb{N}$ and we consider $\gamma : [0,1] \to \mathcal{X}$ be a curve satisfying Assumptions 5.3.8 for $u_0 := u_n$, $u_1 := u_m$ and $y = u$. Using now the inequalities given by Assumptions 5.3.8 at $t = 1/2$, we obtain

$$\begin{aligned}
\mathcal{I} &\leq \phi\big(\gamma(1/2)\big) + \frac{d^2(\gamma(1/2), x)}{2\tau} \\
&\leq \frac{1}{2}\left(\phi(u_n) + \frac{d^2(u_n, u)}{2\tau} + \phi(u_m) + \frac{d^2(u_m, u)}{2\tau}\right) - \frac{1+\lambda\tau}{8\tau}d^2(u_n, u_m)
\end{aligned}$$

Thus, we have

$$\limsup_{n,m\to\infty} \frac{1+\lambda\tau}{8\tau}d^2(u_n, u_m) \leq \limsup_{n,m\to\infty} \frac{1}{2}\left(\phi(u_n) + \frac{d^2(u_n, u)}{2\tau} + \phi(u_m) + \frac{d^2(u_m, u)}{2\tau}\right) - \mathcal{I} = 0.$$

Therefore, the sequence $(u_n)_{n\in\mathbb{N}}$ is Cauchy as soon as $0 < \tau < 1/\lambda$, which shows the uniqueness of discrete solution. On the other hand, thanks to lower semicontinuity of ϕ, we also obtain the existence.
(ii) We claim that the following discrete version of the Evolution Variational Inequality is true, that is for any $u \in \mathcal{X}$ we have that

$$\frac{d^2(u^\tau, v) - d^2(u, v)}{2\tau} + \frac{\lambda}{2}d^2(u^\tau, v) \leq \phi(v) - \phi(u^\tau), \qquad \text{for any } v \in \mathcal{X}, \tag{5.30}$$

where u^τ is the minimizer of (5.12). Indeed, if we pick a curve $\gamma : [0,1] \to \mathcal{X}$, as in Assumptions 5.3.8 for $u_0 := u^\tau, u_1 := v$ and $v := x$, thanks to minimality of u^τ, it holds that

$$\begin{aligned}
\phi(u^\tau) + \frac{d^2(u^\tau, u)}{2\tau} &\leq \phi\big(\gamma(t)\big) \\
&\leq (1-t)\phi(u^\tau) + t\phi(v) - \frac{\lambda}{2}t(1-t)d^2(u^\tau, v) \\
&\quad + \frac{(1-t)d^2(u^\tau, u) + td^2(u, v) - t(1-t)d^2(u^\tau, v)}{2\tau}.
\end{aligned}$$

Thus, taking the limit as $\tau \to 0$, we obtain the desired discrete version of Evolution Variational Inequality, as it was described in (5.30).

Now we will obtain the full convergence of discrete solutions to the limit curve. Given $\overline{u}, \overline{v} \in \overline{D(\phi)}$ and u_t^τ, v_t^τ their discrete solutions respectively, our strategy is to bound the distance $d(u_\tau^{\tau/2}, v_\tau^\tau)$ in terms of distance $d(\overline{u}, \overline{v})$. To do this, we write the discrete version of Evolution Variational

Inequality given by (5.30) two times for $\tau := \tau/2$ and $v := \overline{v}$. The first time with $u := \overline{u}$ and secondly with $u := u^{\tau/2}_{\tau/2}$. Therefore, using the assumption of non-negativeness of λ, we obtain

$$\frac{d^2\big(u^{\tau/2}_{\tau/2}, \overline{v}\big) - d^2(\overline{u}, \overline{v})}{\tau} \leq \phi(\overline{v}) - \phi\big(u^{\tau/2}_{\tau/2}\big) \tag{5.31}$$

$$\frac{d^2(u^{\tau/2}_{\tau}, \overline{v}) - d^2\big(u^{\tau/2}_{\tau/2}, \overline{v}\big)}{\tau} \leq \phi(\overline{v}) - \phi\big(u^{\tau/2}_{\tau}\big). \tag{5.32}$$

Adding up the inequalities (5.31) and (5.32), observing that $\phi(u^{\tau/2}_{\tau}) \leq \phi(u^{\tau/2}_{\tau/2})$, we can see that it holds

$$\frac{d^2(u^{\tau/2}_{\tau}, \overline{v}) - d^2(\overline{u}, \overline{v})}{\tau} \leq 2\big(\phi(\overline{v}) - \phi(u^{\tau/2}_{\tau})\big). \tag{5.33}$$

On the other hand, from discrete version of the Evolution Variational Inequality given by (5.30), written for $u := \overline{v}$ and $v := u^{\tau/2}_{\tau}$, we have

$$\frac{d^2(v^{\tau}_{\tau}, u^{\tau/2}_{\tau}) - d^2(u^{\tau/2}_{\tau}, \overline{v})}{\tau} \leq 2\big(\phi(u^{\tau/2}_{\tau}) - \phi(v^{\tau}_{\tau})\big). \tag{5.34}$$

Again, adding up (5.33) and (5.34), we obtain

$$\frac{d^2(y^{\tau}(\tau), u^{\tau/2}_{\tau}) - d^2(\overline{u}, \overline{y})}{\tau} \leq 2\big(\phi(\overline{y}) - \phi(y^{\tau}_{\tau})\big). \tag{5.35}$$

Now, we pick $t = n\tau$ and $s = m\tau$ with $t < s$, and re-write the discrete version of the Evolution Variational Inequality given by (5.30), for $u := u^{\tau}_{i\tau}$, for $i = n, \cdots, m-1$ and adding everything up to get

$$\frac{d^2(u^{\tau}_t, v) - d^2(u^{\tau}_s, v)}{2(s-t)} + \frac{\lambda\tau}{2(s-t)} \sum_{i=n+1}^{m} d^2(u^{\tau}_{i\tau}, v) \leq \phi(v) - \frac{\tau}{s-t} \sum_{i=n+1}^{m} \phi(u^{\tau}_{i\tau}). \tag{5.36}$$

In the same fashion, we pick $t = n\tau$ and rewrite (5.35) for $\overline{u} := u^{\tau/2}_{i\tau}$ and $\overline{v} := v^{\tau}_{i\tau}$, for $i = 0, \cdot, n-1$. Adding them all up, we obtain

$$\frac{d^2(u^{\tau/2}_t, v^{\tau}_t) - d^2(\overline{u}, \overline{v})}{\tau} \leq 2\big(\phi(\overline{v}) - \phi(v^{\tau}_t)\big)$$

Let $\overline{v} = \overline{u}$, and since $\phi \geq 0$, we obtain

$$d^2(u^{\tau/2}_t, u^{\tau}_t) \leq 2\tau\big(\phi(\overline{u}) - \phi(u^{\tau}_t)\big) \leq 2\tau\phi(\overline{u}). \tag{5.37}$$

Now, we are proceed to a limiting process. For that reason, we consider $\tau = \tau/2^n$ and thus we obtain

$$d^2\big(u^{\tau/2^{n+1}}_t, u^{\tau/2^n}_t\big) \leq \frac{\tau}{2^{n-1}}\phi(\overline{u}).$$

Therefore, we have that

$$d^2(u^{\tau/2^n}_t, u^{\tau/2^m}_t) \leq \tau(2^{2-n} - 2^{2-m})\phi(\overline{u}), \qquad \text{for any } n < m \in \mathbb{N}.$$

This tells us that the sequence $(x^{\tau/2^n}_t)_{n\in\mathbb{N}}$ a is Cauchy sequence for any $t \geq 0$. In addition, setting $n = 0$ and letting $m \to \infty$ we can obtain the discrete version of Evolution Variational Inequality.

In order to achieve the Evolution Variational Inequality, we let $\tau \to 0$ in (5.36), and after calculations, we obtain

$$\frac{d^2(u_t, v) - d^2(u_s, v)}{2(s-t)} + \frac{\lambda}{2(s-t)} \int_t^s d^2(u_r, v)dr \leq \phi(y) - \frac{1}{s-t} \int_t^s \phi(u_r)dr,$$

which is the Evolution Variational Inequality in integral form, so our proof is completed. □

5.4 Gradient flows in Wasserstein space

Recalling the discussion related with Wasserstein spaces in Chapter 4, we have already seen that, given a metric space $(\mathcal{X}, d)$, the space $\mathbb{W}_2(\mathcal{X})$ enjoys several nice metric, topological and geometrical properties. So, given all of these nice and desired properties of Wasserstein spaces, and recalling the discussion of gradient flows in metric setting, one may concern about the possibility to define and study gradient flows in Wasserstein space. Such a concern was the key ingreading of the seminal and revolutionary paper of R. Jordan, D. Kinderlehrer and F. Otto, in late 90s (see Jordan et al. (1998)), which essentially was the begining of study of gradient flows in Wasserstein spaces and the ancestor of series of papers, answering several important related questions (see e.g Otto (2001)).

We will focus in the following programme: at first we will discuss some important facts of subdifferentials of λ-geodesically convex functionals defined on $\mathbb{W}_2(\mathbb{R}^d)$, which in effect comes from the dynamic formulation of optimal transport problem and the underlying Riemannian structure of the Wasserstein space $\mathbb{W}_2(\mathbb{R}^d)$. Secondly, we will present three classical applications, for which we can employ Evolution Variational Inequality formulations, in order to obtain solutions and characterizations of corresponding gradient flows.

5.4.1 Subbdifferential Calculus

As we have see in Chapter 4, Wasserstein spaces enjoy a rich and important geometrical structure. Perhaps, one of the key facts of the importance of this geometrical structure stands to the possibility which gives us to imitating Hilbert spaces and develop an analogous "subdifferential "theory" about λ-geodesically convex functionals defined on $\mathbb{W}_2(\mathbb{R}^d)$. The basic ideas behind of this approach, as we will see, have surprisingly many similarities with the classical subdifferential calculus of convex functionals defined in any separable Hilbert space.

So, following the spirit of their similarities, we will develop the whole theory based on regular measures defined on $\mathbb{R}^d$, keeping the level of technicalities through discussion as simple as possible, presenting the main ideas. Although, paying the costs of (beautiful but pretty complicated) technicalities one can extend the whole theory to more general setting and more general measures.

So, under this perspective, let us begin the exploration of subdifferential calculus in Wasserstein space, by recalling the definition of subdifferential in any separable Hilbert space. Hence, let us recall that $\mathcal{X} = \mathbb{R}^d$, (or even a Hilbert space), and $\phi : \mathcal{X} \to \mathbb{R}$ is a λ-convex functional, then the subdifferential of F at point $x \in \mathcal{X}$ is defined as the set

$$\partial F(x) := \left\{ v \in \mathcal{X} : F(x) + \langle v, y - x \rangle + \frac{\lambda}{2}|x - y|^2 \leq F(y) \quad \text{for every } y \in \mathcal{X} \right\}.$$

Following this spirit, any recalling the λ-convexity of a functional $\phi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$, let us extended analogously the above classical subdifferential definition in Wasserstein space $\mathbb{W}_2(\mathbb{R}^d)$.

Definition 5.4.1 (Subdifferential in $\mathbb{W}_2(\mathbb{R}^d)$)**.** Consider that $\phi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ be a λ-geodesically convex and lower semicontinuous functional, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a regular measure such that $\phi(\mu) < \infty$. If for every $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, denotes as T_μ^ν the optimal transport map between μ and ν, we define the subdifferential of ϕ at μ as

$$\partial^{\mathcal{W}} \phi(\mu) := \left\{ v \in L^2(\mu) : \phi(\mu) + \int_{\mathbb{R}^d} \langle T_\mu^\nu - \mathsf{Id}, v \rangle d\mu + \frac{\lambda}{2} \mathcal{W}_2^2(\mu, \nu) \leq \phi(\nu) \right\}.$$

Let us mention that under this perspective, we have no worries about the existence of such an optimal transport map since that Brenier's theorem (Theorem 3.4.7), guarantees this fact. So, given the notion of subdifferentials in Wasserstein spaces, shall we proceed a closer examination of them. Moreover, let us explore one of their fundamental properties, their *monotonicity,* which imitates the monotonicity property of λ-convex functional in $\mathbb{R}^d$ (or of course, any separable Hilbert space). This property is presented in the following lemma.

Lemma 5.4.2 (Monotonicity of $\partial^{\mathcal{W}}$). *For every $\mu, \nu \in D(\phi)$, $v \in \partial^{\mathcal{W}}\phi(\mu)$ and $w \in \partial^{W}\phi(\nu)$. Then it holds that*

$$\int_{\mathbb{R}^d} \langle v, T_\mu^\nu - \mathsf{Id}\rangle d\mu + \int_{\mathbb{R}^d} \langle w, T_\mu^\nu - \mathsf{Id}\rangle d\nu \leq -\lambda \mathcal{W}_2^2(\mu,\nu).$$

Proof. By the definition of the subdifferential in $\mathbb{W}_2(\mathbb{R}^d)$, we have that

$$\phi(\mu) + \int_{\mathbb{R}^d} \langle T_\mu^\nu - \mathsf{Id}, v\rangle d\mu + \frac{\lambda}{2}\mathcal{W}_2^2(\mu,\nu) \leq \phi(\nu), \tag{5.38}$$

and also

$$\phi(\nu) + \int_{\mathbb{R}^d} \langle T_\mu^\nu - \mathsf{Id}, w\rangle d\nu + \frac{\lambda}{2}\mathcal{W}_2^2(\mu,\nu) \leq \phi(\mu). \tag{5.39}$$

Adding up (5.38) and (5.39), the desired result follows. □

Based on the definition of subdifferential $\partial^{\mathcal{W}}$, we can obtain a first notion of gradient flow in Wasserstein spaces. Again, its definition stands as an analogue of the Hilbert space theory, except one important fact. In such a case, we employ the rich geometrical structure of Wasserstein spaces, which is based on the dynamical formulation optimal transport problem, as it was presented in Section 4.4.

Definition 5.4.3 (Subdifferential Gradient Flow formulation). Consider that $\phi : \mathscr{P}_2(\mathbb{R}^d)$ is a λ-geodesically convex functional and $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a regular measure. Then a locally absolutely curve $(\mu_t)_{t\in[0,1]}$ is a Gradient Flow for ϕ, starting at μ, if $\mu_t \to \mu$ with respect to $\mathcal{W}_2$ and it holds

$$-v_t \in \partial^{\mathcal{W}}\phi(\mu,t), \qquad \text{a.e. for } t>0$$

where $(v_t)_t$ is the vector field which is uniquely identified by the curve $(\mu_t)_{t\in(0,1)}$ via continuity equation

$$\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0, \text{ for every } v_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)) \text{ a.e. for } t>0.$$

On account of the subdifferential formulation, we have by now four ways to study the gradient flows of λ-geodesically convex functionals defined on $\mathscr{P}_2(\mathbb{R}^d)$, which are Energy Dissipation Inequality, Energy Dissipation Equality, Evolution Variational Inequality, and Subdifferential formulation.

The good news are that all of these four formulations are *equivalent* for λ-geodesically convex functionals. This fact is witness by the following proposition.

Proposition 5.4.4 (Equivalence between formulations). *Consider that $\phi : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}\cup\{+\infty\}$ is a λ-geodesically convex functional and $(\mu_t)_{t\in(0,1)}$ is a curve made by regular measures for every $t>0$. Then, the Energy Dissipation Inequality, Energy Dissipation Equality, Evolution Variational Inequality, and Subdifferential formulations are equivalent.*

Proof. Thanks to (iii) of Proposition 4.4.4, we have

$$\frac{1}{2}\frac{d}{dt}\mathcal{W}_2^2(\mu_t,\nu) = -\int_{\mathbb{R}^d}\langle v_t, T_{\mu_t}^\nu\rangle d\mu_t, \text{ a.e. } t>0,$$

where $T_{\mu_t}^{\nu}$ stands for the optimal transport map between μ_t and ν. Then, we have that

$$-v_t \in \partial^{\mathcal{W}}\phi(\mu_t) \Longleftrightarrow \phi(\mu_t) + \int_{\mathbb{R}^d} \langle -v_t, T_{\mu_t}^{\nu} - \mathsf{Id}\rangle d\mu + \frac{\lambda}{2}\mathcal{W}_2^2(\mu_t,\nu) \leq \phi(\nu), \text{ for every } \nu \in \mathscr{P}_2(\mathbb{R}^d),$$

which is equivalent to

$$\phi(\mu_t) + \frac{1}{2}\frac{d}{dt}\mathcal{W}_2^2(\mu_t,\nu) + \frac{\lambda}{2}\mathcal{W}_2^2(\mu_t,\nu) \leq \phi(\nu), \quad \text{ for every } \nu \in \mathscr{P}_2(\mathbb{R}^d) \text{ a.e. } t > 0.$$

□

5.4.2 Three Classical Gradient Flows in Wasserstein spaces

Now we are ready to present three classical examples of gradient flows in Wasserstein spaces. The most suitable way to proceed to such an application, comes from the metric theory of gradient flows for λ-geodesically convex functionals, which was developed in previous section. Moreover, we will study of the associated PDE with the corresponding gradient flow.

To do this we introduce the notion of *interpolating curves* (or McCann interpolants (see McCann (1997))) in the following definition.

Definition 5.4.5 (Interpolating curves)**.** Consider that $\mu, \nu_0, \nu_1 \in \mathscr{P}_2(\mathbb{R}^d)$, where μ is a regular measure. If $\boldsymbol{T}_0, \boldsymbol{T}_1$ are the optimal transport maps between μ and ν_0, ν_1 respectively, then the interpolating curve $(\nu_t)_{t\in(0,1)}$ between ν_0 and ν_1 with base μ is defined as

$$\nu_t := \big((1-t)\boldsymbol{T}_0 + t\boldsymbol{T}_1\big)_{\#}\mu$$

Let us now make some comments on the above definition of interpolating curves. Firstly, it is now hard to see that if for the base measure holds that $\mu = \nu_0$, then the above definition is reduced to the classical definition of geodesic connecting ν_0 to ν_1. In addition, another interesting observation is that, in account of the application of the theory for λ-geodesically convex functionals which was described in Section 5.3, we should define interpolating curves having as base measure *any* measure $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, and not just a regular one. This cab be possible, and all of the forthcoming discussion can be applied in more general settings.

We shall now present an important and useful characterization of interpolating curves.

Proposition 5.4.6. *For any interpolating curve, as it described in Definition 5.4.5, and for any* $t \in [0,1]$ *it holds that*

$$\mathcal{W}_2^2(\mu,\nu_t) \leq (1-t)\mathcal{W}_2^2(\mu,\nu_0) + t\mathcal{W}_2^2(\mu,\nu_1) - t(1-t)\mathcal{W}_2^2(\nu_0,\nu_1).$$

Proof. Since $\boldsymbol{T}_0, \boldsymbol{T}_1$ are optimal transport maps between μ and ν_0, ν_1, due to Brenier's theorem (Theorem 3.4.7), we know that each of them are gradient of a convex function, let us say ϕ_0, ϕ_1 respectively. Moreover, we know the convex combination of them is still convex. Thus, $(1-t)\boldsymbol{T}_0 + t\boldsymbol{T}_1$ is a gradient of $(1-t)\phi_0 + t\phi_1$ and therefore, again thanks to Brenier's theorem, it is still optimal.

In addition, generally speaking, if $f, g : \mathcal{X} \to \mathcal{Y}$ are two Borel measurable maps between metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$, then it holds trivially that

$$\mathcal{W}_2^2(f_{\#}\mu, g_{\#}\mu) \leq \int_{\mathcal{X}} d_{\mathcal{Y}}^2(f(x), g(x))d\mu(x).$$

Employing this observation in our case, we can obtain

$$\mathcal{W}_2^2(\nu_0,\nu_1) \leq \|\boldsymbol{T}_0 - \boldsymbol{T}_1\|_{L^2(\mu)}^2.$$

Moreover, based on this fact we have that

$$\begin{aligned}
\mathcal{W}_2^2(\mu,\nu_t) &= \|(1-t)\boldsymbol{T}_0 + t\boldsymbol{T}_1\|^2_{L^2(\mu)}\\
&= (1-t)|\boldsymbol{T}_0 - \mathsf{Id}\|^2_{L^2(\mu)} + \|\boldsymbol{T}_1 - \mathsf{Id}\|^2_{L^2(\mu)} - t(1-t)|\boldsymbol{T}_0 - \boldsymbol{T}_1\|^2_{L^2(\mu)}\\
&\leq (1-t)\mathcal{W}_2^2(\mu,\nu_0) + T\mathcal{W}_2^2(\mu,\nu_1) - t(1-t)W_2^2(\nu_0,\nu_1),
\end{aligned}$$

which makes our proof complete. □

Now, we are ready to proceed to the definitions of three important functionals, which we will study in the following.

Definition 5.4.7 (Potential energy)**.** Consider that $V:\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$ is a lower semicontinuous and bounded from below functional. The potential energy functional $\mathscr{V}:\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}\cup\{+\infty\}$ associated with V is defined as

$$\mathscr{V}(\mu) := \int_{\mathbb{R}^d} V d\mu.$$

Definition 5.4.8 (Interaction energy)**.** Consider that $W:\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$ is lower semicontinuous, even and bounded from below functional. The interaction energy functional $\mathscr{W}:\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}\cup\{+\infty\}$ associated with W is defined as

$$\mathscr{W}(\mu) := \frac{1}{2}\int W(x_1-x_2)d\mu\otimes\mu(x_1,x_2).$$

Notice that the definition of interaction energy makes sense, non only for even functionals. The assumptions of evenness, comes truly for technical reasons. To be more precise, evenness allow us to replace the quantity $W(x)$ by $W(x)+W(-x)/2$ and keep the value of functional invariant.

Definition 5.4.9 (Internal energy)**.** Let $u:[0,+\infty)\to\mathbb{R}\cup\{+\infty\}$ is a convex and bounded from below functional such that $u(0)=0$ and

$$\liminf_{z\to 0}\frac{u(z)}{z^\alpha} < -\infty, \qquad \text{for } \alpha > \frac{d}{d+2}.$$

Let also $u'(\infty)=\lim_{z\to\infty}\frac{u(z)}{z}$. The internal energy functional $\mathscr{E}$ associated with u is defined as

$$\mathscr{E}(\mu) := \int u(\rho)\mathscr{L}^d + u'(\infty)\mu_s(\mathbb{R}^d),$$

where $\mu=\rho\mathscr{L}^d+\mu_s$ is the Lebesgue decomposition into absolute continuous and singular parts.

Under certain assumptions of $\mathscr{V},\mathscr{W}$ and $\mathscr{E}$, they can be compatible with the distance $\mathcal{W}_2$. Therefore, we can ensure that it is possible to apply the theory which developed in previous section.

The core component of doing this, is nothing more but the definition of interpolating curves. This fact is witnessed by the following proposition.

Proposition 5.4.10. *Consider that $\lambda>0$ and the functionals $\mathscr{V},\mathscr{W}$ and $\mathscr{E}$, as they were defined in Definitions 5.4.7, 5.4.8 and 5.4.9 respectively. Then the following holds:*

(i) The functional $\mathscr{V}$ is λ-convex along then interpolating curves in $\mathbb{W}_2(\mathbb{R}^d)$ if and only if V is λ-convex.

(ii) The functional $\mathscr{W}$ is λ-convex along the interpolating curves in in $\mathbb{W}_2(\mathbb{R}^d)$ if and only if W is λ-convex.

(iii) If the map $z \mapsto z^d u(z^{-d})$ is convex and nondecreasing in $(0, +\infty)$, then the functional $\mathscr{E}$ is convex along interpolating curves in $\mathbb{W}_2(\mathbb{R}^d)$.

Theorem 5.4.11 (General existence result for gradient flows)**.** *Consider that $\mathscr{F}$ is either $\mathscr{V}, \mathscr{W}$ or $\mathscr{E}$ and is also λ-convex along interpolating curves. Then for every $\overline{\mu} \in \mathscr{P}_2(\mathbb{R}^d)$ there exists gradient flow $(\mu_t)_{t\in[0,1]}$ for $\mathscr{F}$ starting from $\overline{\mu}$ in Evolution Variational Inequality sense. Moreover, the curve $(\mu_t)_{t\in[0,1]}$ is locally absolutely continuous, $\mu_t \to \overline{\mu}$, and if μ_t is regular for every $t \in [0,1]$, then it holds*

$$-v_t \in \partial^{\mathcal{W}} \mathscr{F}(\mu_t), \qquad \textit{for a.e. } t \in [0,1] \tag{5.40}$$

where v_t is the velocity field related to μ_t and characterized by

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0$$

$$v_t \in Tan_{\mu_t}\mathscr{P}_2(\mathbb{R}^d) \qquad \textit{for a.e. } t \in [0,1].$$

Remark 5.4.12 (What kind of equation is satisfied?)**.** So far, we have to understand which kind of equations is satisfied by the gradient flow μ_t. As (5.40), witness this is equivalent to compute the subdifferential of each of one the corresponding functionals at an arbitrary point $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. The basic idea behind of this calculations is the following. Under suitable smoothness assumptions, we will focus the equivalence:

$$v \in \partial^{\mathcal{W}} \mathscr{F}(\mu) \overset{smooth}{\Longleftrightarrow} \lim_{\epsilon \to 0} \frac{\mathscr{F}((\mathsf{Id} + \epsilon\nabla\varphi)_{\#}\mu) - \mathscr{F}(\mu)}{\epsilon} = \int_{\mathbb{R}^d} \langle v, \nabla\varphi \rangle d\mu_t, \tag{5.41}$$

for every $\varphi \in C_c^\infty(\mathbb{R}^d)$ is true for any λ-geodesically convex functional. At this point, let us clarify why (5.41) is true. For the implication ($\Leftarrow$), let us fix $\varphi \in C_c^\infty(\mathbb{R}^d)$ and recall that for small enough ϵ, the map $(\mathsf{Id} + \epsilon\nabla\varphi)$ is still optimal (see Corollary 3.4.3). Hence, thanks to definition of subdifferential in Wasserstein spaces, we have

$$\mathscr{F}(\mu) + \epsilon \int_{\mathbb{R}^d} \langle v, \nabla\varphi \rangle d\mu + \epsilon^2 \frac{\lambda}{2} \|\nabla\varphi\|^2_{L^2(\mu_t)} \leq \mathscr{F}((\mathsf{Id} + \epsilon\nabla\varphi)_{\#}\mu).$$

Having this relation on the table, we can subtract $\mathscr{F}(\mu)$ on both sides, divide by ϵ and letting $\epsilon \to 0$. Then the implication follows clearly.

For the implication ($\Rightarrow$), let us pick $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, and let $\boldsymbol{T} : \mathbb{R}^d \to \mathbb{R}^d$ optimal transport map between μ and ν. Thanks to Brenier's theorem (Theorem 3.4.7), we know that $\boldsymbol{T}$ is the gradient of a convex function, let us say ϕ. Under this perspective, we define $\varphi(x) := \phi(x) - |x|^2/2$, and we observe that μ_t can be rewritten as

$$\mu_t = \big((1-t)\mathsf{Id} + t\boldsymbol{T}\big)_{\#}\mu = \big((1-t)\mathsf{Id} + t\nabla\varphi\big)_{\#} = (\mathsf{Id} + t\nabla\varphi)_{\#}\mu.$$

Moreover, upon to λ-convexity assumption on $\mathscr{F}$, we have

$$\mathscr{F}(\nu) \geq \mathscr{F}(\mu) + \frac{d}{dt}\Big|_{t=0} \mathscr{F}(\mu_t) + \frac{\lambda}{2}\mathcal{W}_2(\mu, \nu).$$

Hence, since we have that

$$\frac{d}{dt}\Big|_{t=0} \mathscr{F}(\mu_t) = \int_{\mathbb{R}^d} \langle v, \nabla\varphi \rangle d\mu,$$

thanks to arbitrariness of ν, we conclude that $v \in \partial^{\mathcal{W}} \mathscr{F}(\mu)$.

Proposition 5.4.13 (Gradient flow w.r.t. $\mathscr{V}$)**.** *Consider that $V : \mathbb{R}^d \to \mathbb{R}^d$ is λ-convex and C^1. Let also $\mathscr{V}$ defined according to Definition 5.4.7 and a regular measure $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ such that $\mathscr{V}(\mu) < \infty$. Then $\partial^{\mathcal{W}}$ is non-empty if and only if $\nabla V \in L^2(\mu)$. In such a case, $\nabla V = \partial^{\mathcal{W}}\mathscr{V}(\mu)$. Moreover, if $(\mu_t)_{t\in[0,1]}$ is a gradient flow of $\mathscr{V}$, constructed by regular measures, it solves the equation*

$$\frac{d}{dt}\mu_t = \nabla \cdot (\nabla V \mu_t), \qquad \textit{in distributional sense.}$$

Proof. Let us pick up $\varphi \in C_c^\infty(\mathbb{R}^d)$ and observe that

$$\lim_{\epsilon\to 0} \frac{\mathscr{V}((\mathsf{Id} + \epsilon\nabla\varphi)_{\#}\mu - \mathscr{V}(\mu)}{\epsilon} = \lim_{\epsilon\to 0}\int_{\mathbb{R}^d} \frac{\mathscr{V}\circ(\mathsf{Id}+\epsilon\nabla\varphi) - \mathscr{V}}{\epsilon} d\mu = \int_{\mathbb{R}^d}\langle \nabla\mathscr{V}, \nabla\varphi\rangle d\mu.$$

Thusly, taking into account (5.41), the desired result follows. This fact makes our sketch of proof complete. □

Proposition 5.4.14 (Gradient flow w.r.t. $\mathcal{W}$)**.** *Consider that $W : \mathbb{R}^d \to \mathbb{R}^d$ is λ-convex and C^1. Let also $\mathscr{W}$ defined according to Definition 5.4.9 and a regular measure $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ such that $\mathscr{W}(\mu) < \infty$. Then $\partial^{\mathcal{W}}\mathscr{W}(\mu$ is non-empty if and only if $(\nabla\mathscr{W}) * \mu \in L^2(\mu)$. In such a case $\partial^{\mathcal{W}}\mathscr{W}(\mu) = (\nabla\mathscr{W}) * \mu$. Moreover, if $(\mu_t)_{t\in[0,1]}$ is a gradient flow of $\mathscr{V}$, constructed by regular measures, it solves the non local equation*

$$\frac{d}{dt}\mu_t = \nabla \cdot ((\nabla W * \mu_t)\mu_t), \qquad \textit{in distributional sense.}$$

Proof. Let us pick up $\varphi \in C_c^\infty(\mathbb{R}^d)$ and define $\mu_\epsilon := (\mathsf{Id} + \epsilon\nabla\varphi)$. Then, we observe that

$$\begin{aligned}\mathscr{W}(\mu^\epsilon &= \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)d\mu^\epsilon(x)d\mu^\epsilon(y) = \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y+\epsilon(\nabla\varphi(x)-\nabla\varphi(y)))d\mu(x)d\mu(y)\\ &= \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)d\mu(x)d\mu(y) + \frac{\epsilon}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\langle\nabla W(x-y), \nabla\phi(x) - \nabla(y)\rangle d\mu(x)d\mu(y) + o(\epsilon).\end{aligned}$$

Moreover, we observe that

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d}\langle\nabla W(x-y), \nabla\varphi(x)\rangle d\mu(x)d\mu(y) &= \int_{\mathbb{R}^d}\Big\langle\int_{\mathbb{R}^d}\nabla W(x-y)d\mu(y), \nabla\varphi(x)\Big\rangle d\mu(x)\\ &= \int_{\mathbb{R}^d}\langle\nabla W * \mu(x), \nabla\varphi(x)\rangle d\mu(x).\end{aligned}$$

In the same fashion, we can obtain

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\langle\nabla W(x-y), \nabla\varphi(y)\rangle d\mu(x)d\mu(y) = \int_{\mathbb{R}^d}\langle\nabla W*\mu(y), \nabla\varphi(y)\rangle d\mu(y) = \int_{\mathbb{R}^d}\langle\nabla W*\mu(x), \nabla\varphi(x)\rangle d\mu(x).$$

So, employing (5.41), the desired result follows, and concludes our sketch of proof. □

Proposition 5.4.15 (Gradient flow w.r.t. $\mathcal{E}$)**.** *Consider that $u : [0, +\infty) \to \mathbb{R}$ is a convex C^2 and bounded from below functional satisfying equations A,B. Let also $\mu = \rho\mathscr{L}^d \in \mathscr{P}_2(\mathbb{R}^d)$ be an absolutely continuous measure with smooth density. Then $\partial^{\mathcal{W}}\mathscr{E}(\mu) = \nabla(u'(\rho))$. Moreover, if μ_t is a gradient flow for $\mathscr{E}$ and absolutely continuous with smooth density ρ_t for every $t > 0$, then the map $t \mapsto \rho_t$ solves the equations*

$$\frac{d}{dt}\rho_t = \nabla\cdot(\rho_t\nabla(u'(\rho_t))).$$

Proof. Let us pick $\varphi \in C_c^\infty(\mathbb{R}^d)$ and define $\mu^\epsilon := (\mathsf{Id} + \epsilon\nabla\varphi)_{\#}\mu$. Then for small enough ϵ, μ_ϵ is absolutely continuous. Moreover, up to a change of variable formula, its density ρ^ϵ satisfies the formula

$$\rho^\epsilon(x + \epsilon\nabla\varphi(x)) = \frac{\rho(x)}{\det(\mathsf{Id} + \epsilon\nabla^2\varphi(x))}.$$

Then, using the fact that

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0}(\det(\mathsf{Id} + \epsilon\nabla^2\varphi(x))) = \Delta\varphi(x),$$

we can obtain

$$\begin{aligned}
\frac{d}{d\epsilon}\bigg|_{\epsilon=0}\mathscr{E}(\mu^\epsilon) &= \frac{d}{d\epsilon}\bigg|_{\epsilon=0}\int_{\mathbb{R}^d} u(\rho^\epsilon(y))dy \\
&= \frac{d}{d\epsilon}\bigg|_{\epsilon=0}\int_{\mathbb{R}^d} u\Big(\frac{\rho(x)}{\det(\mathsf{Id} + \epsilon\nabla^2\varphi(x))}\Big)\det(\mathsf{Id} + \epsilon\nabla^2\varphi(x)))dx \\
&= \int_{\mathbb{R}^d} -\rho'(\rho)\Delta\varphi + u(\rho)\Delta\varphi dx = \int_{\mathbb{R}^d}\langle\nabla(\rho u'(\rho) - u(\rho)), \nabla\varphi\rangle dx \\
&= \int_{\mathbb{R}^d}\langle\nabla(u'(\rho)), \nabla\varphi\rangle\rho dx.
\end{aligned}$$

Under this considerations, the desired result follows from equivalence (5.41), which makes our sketch of proof complete. □

Bibliography

Ambrosio, L. , Gigli N. (2013). A user's guide to Optimal Transport. In *Modelling and optimisation of flows on networks* (pp. 1–155). Springer.

Ambrosio, L. (1995). Minimizing Movements. *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl.(5)*, *19*, 191–246.

Ambrosio, L., Gigli, N., & Savaré, G. (2008). *Gradient Flows: in metric spaces and in the space of Probability Measures.* Springer Science & Business Media.

Ambrosio, L., Kirchheim, B., Pratelli, A., et al. (2004). Existence of optimal transport maps for crystalline norms. *Duke Mathematical Journal*, *125*(2), 207–241.

Appell, P. (1887). Mémoire sur les déblais et les remblais des systemes continus ou discontinus. *Mémoires présentes par divers Savants à l'Académie des Sciences de l'Institut de France*, *29*, 1–208.

Aliprantis, C.D., & Border, K. C. (2006). *Infinite Dimensional Analysis: a hitchhiker's guide.* Springer.

Bachman, G., & Narici, L. (2000). Functional Analysis, rev. ed. *Mineola, NY: Dover*.

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

Benyamini, Y., & Lindenstrauss, J. (1998). *Geometric Nonlinear Functional Analysis* (Vol. 48). American Mathematical Soc.

Billingsley, P. (2013). *Convergence of probability measures.* John Wiley & Sons.

Borwein, J. M., Vanderwerff, J. D., et al. (2010). *Convex Functions: Constructions, Characterizations and Counterexamples* (Vol. 109). Cambridge University Press Cambridge.

Brasco, L. (2012). A survey on dynamical transport distances. *Journal of Mathematical Sciences*, *181*(6), 755–781.

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, *44*(4), 375–417.

Brézis, H. (1971). Monotonicity methods in hilbert spaces and some applications to nonlinear partial differential equations. In *Contributions to nonlinear functional analysis* (pp. 101–156). Elsevier.

Brezis, H., & Browder, F. (1998). Partial differential equations in the 20th century. *Advances in Mathematics*, *135*(1), 76–144.

Burago, D., Burago, Y., Burago, I. D., & Ivanov, S. (2001). *A Course in Metric Geometry* (Vol. 33). American Mathematical Soc.

Cardinali, T., Colombo, G., Papalini, F., & Tosques, M. (1997). On a class of evolution equations without convexity. *Nonlinear analysis*, *28*(2), 217–234.

Carmo, M. P. d. (1992). *Riemannian Geometry.* Birkhäuser.

Carrillo, J. A., McCann, R. J., & Villani, C. (2006). Contractions in the 2-wasserstein length space and thermalization of granular media. *Archive for Rational Mechanics and Analysis*, *179*(2), 217–263.

Cayley. (1883). On monge's "mémoire sur la théorie des déblais et des remblais.". *Proceedings of the London Mathematical Society*, *s1-14*(1), 139–143.

Champion, T., De Pascale, L., & Juutinen, P. (2008). The ∞-Wasserstein distance: Local solutions and existence of optimal transport maps. *SIAM Journal on Mathematical Analysis*, *40*(1), 1–20.

Crandall, M., & Liggett, T. (1971). Generation of semi-groups of nonlinear transformations on general banach spaces. *American Journal of Mathematics*, *93*(2), 265–298.

Darboux, G. (1885). Prix bordin (géométrie). *CR Acad. Sci. Paris*, *101*, 1312–1316.

De Giorgi, E. (1993). New problems on minimizing movements. *Ennio de Giorgi: Selected Papers*, 699–713.

De Giorgi, E., Degiovanni, M., Marino, A., & Tosques, M. (1983). Evolution equations for a class of nonlinear operators. *Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti Lincei. Matematica e Applicazioni*, *75*(1-2), 1–8.

De Giorgi E., M. A., & Tosques, M. (1980). Problems of evolution in metric spaces and maximal decreasing curve. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.(8)*, *68*(3), 180–187.

De Philippis, G., & Figalli, A. (2014). The Monge–Ampère equation and its link to optimal transportation. *Bulletin of the American Mathematical Society*, *51*(4), 527–580.

Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, *15*(3), 458–486.

Fabian, M., Habala, P., Hájek, P., Montesinos, V., & Zizler, V. (2011). *Banach Space Theory: the basis for linear and nonlinear analysis.* Springer Science & Business Media.

Figalli, A. (2017). *The Monge–Ampère equation and its applications* Zurich Lectures in Advanced Mathematics EMS

Gangbo, W., & McCann, R. J. (1996). The Geometry of Optimal Transportation. *Acta Mathematica*, *177*(2), 113–161.

Gangbo, W., Nguyen, T., Tudorascu, A. (2008). Hamilton-Jacobi equations in the Wasserstein space. *Methods and Applications of Analysis*, *15*(2), 155–184.

Garrido, M. I., & Jaramillo, J. A. (2008). Lipschitz-type functions on metric spaces. *Journal of Mathematical Analysis and Applications*, *340*(1), 282–290.

Gianazza, U., & Savaré, G. (1996). Abstract evolution equations on variable domains: an approach by minimizing movements. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, *23*(1), 149–178.

Gromov, M. (2007). *Metric structures for Riemannian and Non-Riemannian spaces.* Springer Science & Business Media.

Hadamard, J. (1923). *Lectures on Cauchy's problem in linear Partial Differential Equations* (Vol. 37). Yale University Press.

Hiriart-Urruty, J.-B., & Lemaréchal, C. (2012). *Fundamentals of Convex Analysis.* Springer Science & Business Media.

Jensen, R. (1993). Uniqueness of Lipschitz extensions: minimizing the sup norm of the gradient. *Archive for Rational Mechanics and Analysis*, *123*(1), 51–74.

Jordan, R., Kinderlehrer, D., & Otto, F. (1998). The variational formulation of the Fokker–Planck equation. *SIAM journal on Mathematical Analysis*, *29*(1), 1–17.

Kantorovich, L. V. (1942). On the translocation of masses. In *Dokl. akad. nauk. ussr (ns)* (Vol. 37, pp. 199–201).

Kechris, A. (2012). *Classical Descriptive Set Theory* (Vol. 156). Springer Science & Business Media.

Klainerman, S. (2010). PDE as a unified subject. In *Visions in mathematics* (pp. 279–315). Springer.

Lions, J. (1969). Quelques méthodes de résolution des problemes aux limites non linéaires.

Lions, J., & Stampacchia G. (1967). Variational inequalities. *Communications on pure and applied mathematics*, *20*(3), 493–519.

Lisini, S. (2007). Characterization of absolutely continuous curves in Wasserstein spaces. *Calculus of Variations and Partial Differential Equations*, *28*(1), 85–120.

McCann, R. J. (1997). A convexity principle for interacting gases. *Advances in mathematics*, *128*(1), 153–179.

Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*.

Otto, F. (2001). The geometry of dissipative evolution equations: the porous medium equation.

Papadopoulos, A. (2005). *Metric spaces, convexity and nonpositive curvature* (Vol. 6). European Mathematical Society.

Parthasarathy, K. R. (2005). *Probability measures on metric spaces* (Vol. 352). American Mathematical Soc.

Peyré, G., & Cuturi, M. (2018, Mar). Computational Optimal Transport. *arXiv e-prints*, arXiv:1803.00567.

Phelps, R. R. (2001). *Lectures on Choquet's Theorem.* Springer Science & Business Media.

Rachev, S. T., Klebanov, L., Stoyanov, S. V., & Fabozzi, F. (2013). *The methods of distances in the theory of probability and statistics.* Springer Science & Business Media.

Rockafellar, R. (2015). *Convex Analysis.* Princeton University Press.

Rulla, J. (1996). Error analysis for implicit approximations to solutions to cauchy problems. *SIAM journal on numerical analysis*, *33*(1), 68–87.

Santambrogio, F. (2015). Optimal Transport for Applied Mathematicians. *Birkäuser, NY*, 99–102.

Schwartz, J. T., Schwartz, L., & Karcher, H. (1969). *Nonlinear Functional Analysis* (Vol. 4). CRC Press.

Smith, C. S., & Knott, M. (1987). Note on the optimal transportation of distributions. *Journal of Optimization Theory and Applications*, *52*(2), 323–329.

Topsoe, F. (2006). *Topology and Measure* (Vol. 133). Springer.

Vaserstein, L. N. (1969). Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, *5*(3), 64–72.

Vershik, A. M. (2013). Long history of the Monge-Kantorovich transportation problem. *The Mathematical Intelligencer*, *35*(4), 1–9.

Villani, C. (2003). *Topics in Optimal Transportation* (No. 58). Graduate Studies in MAthematics, AMS

Villani, C. (2009). *Optimal Transport: Old and New* Grundlehren der mathematischen Wissenschaften, Springer

Zajíček, L. (1979). On the differentiation of convex functions in finite and infinite dimensional spaces. *Czechoslovak Mathematical Journal*, *29*(3), 340–348.