2016

Well-defined Lagrangian flows for absolutely continuous curves of probabilities on the line

Mohamed H. Amsaad

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
Amsaad, Mohamed H., "Well-defined Lagrangian flows for absolutely continuous curves of probabilities on the line" (2016). Graduate Theses, Dissertations, and Problem Reports. 5098.
https://researchrepository.wvu.edu/etd/5098

This Dissertation is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Graduate Theses, Dissertations, and Problem Reports by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.
Well-defined Lagrangian flows for absolutely continuous curves of probabilities on the line

Mohamed H. Amsaad

Dissertation submitted
to the Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Mathematics

Adrian Tudorascu, Ph.D., Chair
Harumi Hattori, Ph.D.
Harry Gingold, Ph.D.
Tudor Stanescu, Ph.D.
Charis Tsikkou, Ph.D.

Department of Mathematics

Morgantown, West Virginia
2016

Keywords: Continuity Equation, Lagrangian Flow, Optimal Transport, Wasserstein metric, Wasserstein space.

Copyright © 2016 Mohamed H. Amsaad
ABSTRACT

Well-defined Lagrangian flows for absolutely continuous curves of probabilities on the line

Mohamed H Amsaad

Doctor Of Philosophy in Mathematics

It is known from Fluid Mechanics [25], [39], [57], [62], [66], [75], [83], [84], [90], [91], [93], [105], [108], that the time-evolution of a probability measure describing some physical quantity (such as the density of a fluid) is related to the velocity of the fluid by the continuity equation. This is known as the Eulerian description of fluid flow. Dually, the Lagrangian description uses the flow of the velocity field to look at the individual trajectories of particles. In the case of flows on the real line, only recently has it been discovered [70], [92] that `some sort` of dual Lagrangian flow consisting of monotone maps is always available to match an Eulerian flow. The uniqueness of this `monotone flow` among all possible `flows` (quotation marks used precisely because traditionally it cannot be called a `flow` unless it is unique) of the fluid velocity which is the centerpiece of this dissertation.

The Lagrangian description of absolutely continuous curves of probability measures on the real line is analyzed in this thesis. Whereas each such curve admits a Lagrangian description as a well-defined flow of its velocity field, further conditions on the curve and/or its velocity are necessary for uniqueness. We identify two seemingly unrelated such conditions that ensure that the only flow map associated to the curve consists of a time-independent rearrangement of the generalized inverses of the cumulative distribution functions of the measures on the curve. At the same time, our methods of proof yield uniqueness within a certain class for the curve associated to a given velocity; that is, they provide uniqueness for the solution of the continuity equation within a certain class of curves. Our proposed approach is based on the connection between the flow equation and one-dimensional Optimal Transport.

This is based on joint work of the author with A. Tudorascu [21], in which some results on well-posedness (in the one-dimensional case) have been achieved. The results are presented in major conferences and published in a highly ranked mathematical journal.
Well-defined Lagrangian flows for absolutely continuous curves of probabilities on the line

by

MOHAMED H AMSAAD

Dissertation submitted to the
EBERLY COLLEGE OF ARTS AND SCIENCES
at West Virginia University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
MATHEMATICS

DEPARTMENT OF MATHEMATICS

APPROVAL OF THE EXAMINING COMMITTEE

[Signatures]

HARUMI HATTORI, Ph.D.

HARRY GINGOLD, Ph.D.

TUDOR STANESCU, Ph.D.

CHARIS TSIKKOU, Ph.D.

ADRIAN TUDORASCU, Ph.D., Chair

04/25/2016
Date
Dedication

I'd like to dedicate this work.

To my optimal parents Hassan Amsaad and Mabruka Omran for their immeasurable support over the years during all my life.

To my beloved wife Fatma Naser for taking care of our kids while I was finishing this dissertation.

To my children Hassan, Raghad and Jihad, whose love made my desire to finish my PhD every day and night I stayed away from them.

To my deer siblings Somia, Fathi, Khaled, Naima, Nader, Jamal, Ahlam and Aisha.

To all of my teachers, colleagues and friends for all of their support and encouragement during studies in school and college.

I want further to dedicate this thesis to my grandfather Mohamed Amsaad.
Acknowledgments

First and foremost, I would like to express my gratitude to Allah (God) Almighty for giving me the strength, knowledge, ability and time to undertake this research study and to persevere and complete it satisfactorily. Without his blessings, this achievement would not have been possible. I would then like to thank Dr. Adrian D. Tudorascu, for giving me the opportunity to work with him. He constantly supervised my work and gave a lot of suggestions regarding my research.

I would also like to thank Dr. Harumi Hattori, Dr. Harry Gingold, Dr. Tudor Stanescu and Dr. Charis Tsikkou for being members on my committee. I have been very fortunate to have had the opportunity to take courses with some of them, and their teaching has been essential to my understanding of the subject. I would like to thank the Department of Mathematics of WVU for its support regarding the results obtained in this dissertation. I would like to extend my sincere appreciation to Dr. Harvey Diamond and Dr. Casian Pantea for their great help over my study duration. I wish to thank everybody with whom I had stimulating discussions that led to the inception of this thesis, and their suggestions for indicating ways that led to clarify the presentation, concepts and results.

Last but not least, I would like to express my gratitude to my family for their constant guidance and support. First to my beloved wife and my children who care about my successes for pushing me forward through all the stresses and deadlines, while also dealing with their own. I would be remiss if I did not also express my especially grateful to my parents, brothers, and sisters for giving me the motivation to pursue higher studies and gave me a lot of valuable suggestions which are helpful during my stay here in US. I want to say thanks to all of my professors, colleagues and friends who support me all the time and made my stay at WVU so memorable. Their support seemingly has no limit and has been much appreciated throughout my life. Finally, I gratefully acknowledge support from CBIE grant No. 1515 as well as funds from the State of WV and Libyan Ministry of Higher Education.
Contents

Approval Page 2
Acknowledgments v
List of Figures ix
Notation x
Preface 1

1 An overview on Lagrangian and Eulerian descriptions of fluid flows 16
  1.1 Lagrangian vs Eulerian points of view .......................... 17
  1.2 The classical theory of Lagrangian flows ....................... 20
  1.3 The Continuity Equation in the smooth setting ................... 27
    1.3.1 Weak and Measure–Valued Solutions ....................... 28
    1.3.2 Existence and uniqueness of smooth solutions ............... 31
    1.3.3 Connection to the Transport Equation ..................... 36
  1.4 The Continuity Equation out of the smooth setting ............ 43
    1.4.1 Existence of weak solutions of the continuity equation .... 44
    1.4.2 Vector fields with Sobolev spatial regularity .............. 45
    1.4.3 Vector fields with BV spatial regularity .................. 53

2 Theory of regular Lagrangian flows 55
  2.1 The probabilistic point of view ................................ 56
    2.1.1 Probabilistic solutions of the continuity equation ......... 57
    2.1.2 Probabilistic representation of the measure–valued solutions 60
  2.2 ODE uniqueness vs PDE uniqueness ............................ 61
  2.3 Ambrosio’s notion of Lagrangian flows ......................... 63
    2.3.1 The regular Lagrangian flow ................................ 63
    2.3.2 $L_v$–Lagrangian flows .................................... 64
    2.3.3 Well–posedness of Lagrangian flows ....................... 67
  2.4 DiPerna–Lions’ notion of Lagrangian flows ..................... 71

3 Theory of optimal transportations 73
  3.1 Transport problem in the classical framework .................. 74
  3.2 Kantorovich relaxation of Monge problem ....................... 77
3.2.1 Statement of the primal optimization problem .......................... 78
3.2.2 Existence of optimal transport plans and couplings .................. 83
3.2.3 Dual optimization problem ........................................... 88
3.3 Regularity of Brenier’s map and its applications ...................... 95
3.3.1 Existence and uniqueness of Brenier’s maps .......................... 96
3.3.2 Monge–Ampère equation and regularity ............................. 102
3.3.3 Brenier’s polar factorization and rearrangements .................. 104

4 Unidimensional optimal transports ........................................ 109
4.1 Monotonicity of optimal transports on the real line .................... 110
4.1.1 Inverse distribution functions ....................................... 110
4.1.2 The optimality of transport plans for the quadratic cost .......... 123
4.1.3 Monotone transport maps and plans on $\mathbb{R}$ .................... 126
4.1.4 The optimality of monotone maps for a convex cost ............... 131
4.2 Monge–Kantorovich metrics on the real line ........................... 137
4.2.1 The $p$th–Wasserstein space and $L^p$–Wasserstein distance ..... 138
4.2.2 Topology of the $p$th–Wasserstein space ............................ 142

5 Rearrangements and well-defined of the Lagrangian flows ............. 147
5.1 Lagrangian and Eulerian viewpoints in the 1D case ..................... 148
5.1.1 Lagrangian flows of Borel velocity fields on the real line ....... 148
5.1.2 The unidimensional continuity equation ............................ 150
5.1.3 A strategy to obtain uniqueness of the Lagrangian flows ....... 154
5.1.4 The density of Lagrangian flows on the real line ................. 155
5.2 Lagrangian flows in the $H^1(0,T;L^2(I))$ space ..................... 157
5.2.1 Optimal flows associated to $AC^2(0,T;\mathcal{P}_2(\mathbb{R}))$ curves .. 157
5.2.2 Monotone rearrangement of the optimal flows .................... 159
5.2.3 Examples of a monotonicity and a nonmonotonicity of flows .. 162
5.2.4 Connection between $H^1(0,T;L^2(I))$ and $AC^2(0,T;\mathcal{P}_2(\mathbb{R}))$ spaces .... 169
5.2.5 Uniqueness of the flow associated to $AC^2(0,T;\mathcal{P}_2(\mathbb{R}))$ curves .. 171

6 Lagrangian flows associated to $AC^q(0,T;\mathcal{P}_{ac}^p(\mathbb{R}))$ curves .... 176
6.1 Velocity fields associated to $AC^q(0,T;\mathcal{P}_p(\mathbb{R}))$ curves .......................... 177
6.1.1 Theory of $L^q$–absolutely continuous curves in $\mathcal{P}_p(\mathbb{R})$ ...... 178
6.1.2 Existence and uniqueness of the velocity fields on $\mathbb{R}$ ....... 180
6.1.3 The $L^1$–velocity of absolutely continuous curves in $\mathcal{P}_p(\mathbb{R})$ .... 182
6.2 Lagrangian flows in the $W^{1,q}(0,T;L^p(I))$ space .......................... 185
6.2.1 The spaces $AC^q(0,T;L^p(I))$ and $W^{1,q}(0,T;L^p(I))$ .......... 186
6.2.2 The spaces $W^{1,q}(0,T;L^p(I))$ and $AC^q(0,T;\mathcal{P}_p(\mathbb{R}))$ .......... 192
6.3 Optimal flow maps in the $W^{1,q}(0,T;L^p(I))$ space .................. 195
6.3.1 The spaces $AC^q(0,T;L^p(I))$ and $AC^q(0,T;\mathcal{P}_p(\mathbb{R}))$ .......... 196
6.3.2 Optimal flow maps associated to $AC^q(0,T;\mathcal{P}_{ac}^p(\mathbb{R}))$ curves .. 203
6.3.3 Optimal flow maps associated to $AC^q(0,T;\mathcal{P}_p(\mathbb{R}))$ curves .. 209
CONTENTS

7 Uniqueness for the Lagrangian flow and for the Continuity Equation 213
7.1 Continuous Case ........................................... 214
7.1.1 Regularity of the cumulative distribution function .......... 215
7.1.2 Lagrangian description of jointly continuous curves .......... 219
7.1.3 The $L^1$–velocity of jointly continuous curves in $AC^1(0, T; \mathcal{P}_{ac}^1(\mathbb{R}))$ . . 225
7.1.4 Uniqueness of solutions for the continuity equation .......... 231
7.2 Case of higher integrability .................................. 233
7.2.1 The functional derivative of optimal maps in $L^q(0, T; L^p(I))$ ...... 234
7.2.2 Uniqueness of the Lagrangian flows in $W^{1,1}(0, T; L^1(\mathbb{R}))$ .......... 236

8 Open problems and future work 245

A A background for technical materials 248
A.1 Ingredients from measure theory ................................ 248
A.2 Differentiability of Lipschitz continuous functions .......... 251
A.3 Functions with bounded variation ............................ 252
A.4 Maximal functions ........................................ 253
A.5 Vector–valued maps in Banach Spaces ....................... 255
  A.5.1 Measurability ........................................ 255
  A.5.2 Bochner and weak-∗integral ................................ 256
  A.5.3 Differentiability ....................................... 257
  A.5.4 Absolutely continuous curves in metric spaces .......... 259
  A.5.5 The Radon-Nikodým property ................................ 260

B Geometrical properties for transports 262
B.1 Ingredients from convex analysis ............................. 262
B.2 Monotone functions in $\mathbb{R}^d$ ................................ 265
B.3 Cyclical monotonicity, $c$-convexity and $c$-concavity .......... 267
B.4 Differentiability in the $p^\text{th}$-Wasserstein space .......... 270

References 271
List of Figures

5.1 Flow is satisfied in the integral sense for Example 5.12. .......................... 165
5.2 Flow is satisfied in the classical sense for Example 5.12. .......................... 166
5.3 Flow has different monotone solutions ......................................................... 169
Notation

We use the following notation and symbols throughout this thesis.

\( \mathcal{P}(\mathcal{X}) \) : Probability measures in a separable metric space \( \mathcal{X} \)
\( \mathcal{P}_p(\mathcal{X}) \) : Probability measures with finite \( p \)th–moments
\( \mathcal{P}^{ac}_p(\mathcal{X}) \) : Absolutely continuous probabilities with respect to a specified measure
\( \mathcal{M}_+(\mathcal{X}) \) : Positive locally finite (Borel) measures on \( \mathcal{X} \)
\( T_c(\mu, \nu) \) : Optimal (total) cost between \( \mu, \nu \), and depends on a cost function \( c(x,y) \)
\( \Gamma(\mu, \nu) \) : Set of 2–transport plans with given marginals \( \mu, \nu \), see (3.8)
\( W_p(\mu, \nu) \) : The \( p \)th–Wasserstein distance between \( \mu \) and \( \nu \), see (6.16)
\( \Gamma_o(\mu, \nu) \) : Set of optimal transport plans from \( \mu \) to \( \nu \), see Remark 3.4
\( \text{Id} \) : Identity map on \( \mathcal{X} \)
\( C(\mathcal{X}) \) : Space of continuous real functions defined on \( \mathcal{X} \)
\( C^k_b(\mathcal{X}) \) : Space of \( k \)–times continuously differentiable functions that are bounded
\( C^k_c(\mathcal{X}) \) : The \( k \)–times continuously differentiable functions with compact support
\( \mathcal{D}'(\mathcal{X}), C'(\mathcal{X}) \) : Space of distributions on \( \mathcal{X} \), space of continuous linear functionals on \( \mathcal{X} \)
\( B(\mathcal{X}) \) : Borel sets in a separable metric space \( \mathcal{X} \)
\( \text{Lip}(\psi, A) \) : Lipschitz constant of the function \( \psi \) in the set \( A \)
\( L^p(\mu; \mathcal{X}) \) : \( L^p \) space of \( \mu \)–measurable \( \mathcal{X} \)–valued maps with finite \( L^p \)–norms
\( L^q(a, b; \mathcal{S}) \) : \( L^q \) space of vector-valued functions in a complete metric space \( \mathcal{S} \)
\( |\sigma'|(t) \) : Metric derivative of \( \sigma : (a, b) \to \mathcal{S} \), see Section 1.1 of [17]
\( AC^q(a, b, \mathcal{S}) \) : Absolutely continuous \( \sigma : (a, b) \to \mathcal{S} \) with \( |\sigma'| \in L^q(a, b) \)
\( W^{1,q}(a, b; \mathcal{S}) \) : Metric Sobolev spaces. If \( q = 2 \) this is a Hilbert space \( H^1(a, b; \mathcal{S}) \)
\( \mathbb{R}^d, \mathcal{L}^d \) : The \( d \)–dimensional Euclidean space, the Lebesgue measure on \( \mathbb{R}^d \) for \( d \geq 1 \)
\( |\cdot|, \langle \cdot, \cdot \rangle \) : Euclidean norm on \( \mathbb{R}^d \), standard inner product
\( \omega, \Omega, \xi, \varphi, \ldots \) : Usually a modulus of continuity, a domain in \( \mathbb{R}^d \), and often a test function
\[ |A|, \mathcal{L}^d(A) \] : Lebesgue measure of a set \( A \subset \mathbb{R}^d \)

\[ \psi^* \] : Legendre transform of a function \( \psi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\pm \infty\} \)

\[ B_R(x) \] : Open ball of radius \( R \) centered at \( x \) in a metric space

\[ C_c^\infty(\mathbb{R}^d) \] : Space of smooth real functions with compact support in \( \mathbb{R}^d \)

\[ \mathcal{Mon}(S) \] : Set of monotone rearrangements of functions \( S : (a, b) \rightarrow \mathcal{S} \)

\[ \mathbb{E}[Z] \] : Expectation value of a random variable \( Z \)

\[ \text{Dom}(\psi) \] : Domain of the functional \( \psi \), see Definition 3.14

\[ \delta_a \] : Dirac mass concentrated at point \( a \)

\[ \mathbb{I}_A \] : Indicator function of a set \( A \), equal to 1 on \( A \) and 0 on \( A^c \)

\[ \mathcal{H}^k \] : The \( k \)-dimensional Hausdorff measure

\[ \mu \ll \nu \] : Measure \( \mu \) is absolutely continuous w.r.t. \( \nu \)

\[ f \cdot \mu \] : Measure with density \( f \) w.r.t. \( \mu \)

\[ \mu \mathcal{L} A, \mu|_A \] : Measure \( \mu \) restricted to a set \( A \) (i.e., \( \mathbb{I}_A \cdot \mu \))

\[ f|_A \] : Restriction of a function \( f \) to a set \( A \)

\[ \chi \] : Restriction of \( \mathcal{L}^1 \) to the unit open interval in \( \mathbb{R} \)

\[ S_{\#\mu} \] : Image measure of \( \mu \) through the map \( S \)

\[ \text{Tr} \] : Trace operator on matrices

\[ \mu \otimes \nu \] : Product measure of \( \mu \) and \( \nu \) (s.t. \( \mu \otimes \nu(A \times B) = \mu(A) \nu(B) \))

\[ \gamma_T \] : Transport plan in \( \Gamma(\mu, \nu) \) associated to a map \( T \) with \( T_{\#}\mu = \nu \)

\[ M_c(T), K_c(\gamma) \] : Monge cost of a map \( T \) and Kantorovich cost of a plan \( \gamma \)

\[ \pi_1, \pi_2 \] : The projection of a product space \( X_1 \times X_2 \) onto its components

\[ \Gamma_T \] : Space of continuous curves, parametrized on \([0, T]\) and valued in \( \mathbb{R}^d \)

\[ e_t \] : Evaluation map \( e_t : \Gamma_T \rightarrow \mathbb{R}^d \) at time \( t \); i.e., \( e_t(u(\cdot)) = u(t) \)

\[ \rho_t, \mu_t, v_t, X_t, \ldots \] : Subscript \( t \) denotes the value at time \( t \), not a derivative in time

\[ \partial_t, \frac{d}{dt}, \partial_x, \frac{d}{dx}, \ldots \] : Denote to derivatives in time and in space, respectively

\[ \mathbf{n} \] : Outward normal vector to a given domain

\[ \nabla, \nabla \cdot \] : Denote gradients and divergence (or, \( D \) and \( \text{div} \)), respectively

\[ \Delta \] : Denotes the Laplacian: \( \Delta f = \nabla^2 f = \nabla \cdot (\nabla f) \)

\[ \wedge, \vee \] : The min and max operators: \( a \wedge b := \min\{a, b\}, a \vee b := \max\{a, b\} \)

\[ G(\mathcal{X}) \] : Set of measure-preserving maps on a measure space \( \mathcal{X} \)

\[ F, M, g \] : CDF, its generalized, and measure-preserving map

\[ \mathbf{v}, v, \rho \] : Typically denote to velocity fields in multi-d, in 1D, and the density

Bold letters denote vectors.
The rate of change of a certain quantity (concentration of salt in brine, population growth, temperature evolution in a certain environment) does in many situations depend on the amount of said quantity at the time the rate of change is being measured. Mathematically, the rate of change measured at a certain instant is nothing but the derivative of the function representing the changing quantity. By using generally accepted/verified principles governing the change of a certain entity within a system, people have arrived to the concept of “Differential Equations” (DE’s). Nowadays, it would be inconceivable in science to study the quantitative and qualitative behavior of any changing entity without modeling the process in terms of a system of DE’s. DE’s are largely used in Physics, Chemistry, Biology, Finance, Engineering etc. A very basic, undergraduate textbook example is the growth of unsophisticated populations (say, bacteria) while the environment does not exhibit significant changes. The underlying reasonable assumption in this case is that the rate of change at a certain time in the number of individuals is proportional to the actual number of individuals at that time. Thus, we get to the DE \( \dot{P}(t) = \kappa P(t) \), where \( P(t) \) is the population size at time \( t \geq 0 \) and \( \kappa > 0 \) is the rate of growth. This can be measured experimentally, by counting the individuals at different moments in time. The solution is \( P(t) = P(0)e^{\kappa t} \), and one may now use this to predict the size over time. It should also be noted that the same initial population \( P(0) \) cannot give rise to different later numbers if the growth rate is the same (uniqueness of solution). Also, one may now use the solution \( P(t) \) to understand what happens as time becomes ever larger (asymptotic behavior of solution). In some complex cases, the explicit solution may not be easy (or even possible) to find. However, by analyzing the DE itself (rather than its unknown solution), one may be able to confirm that the equation has solutions (existence of a solution), whether they are unique, and what kind of asymptotic behavior the solutions display. The concept of stability is also a very important one; if one expects (by, say, observation) that small perturbations in the initial data (e.g., \( P(0) \) from above) should result in small differences between solutions at all later times, then the model used should account for that. This concept is non-vacuous provided that solutions exist not only for some particular initial data \( P(0) \), but for sufficiently many different initial data close to \( P(0) \).

The classical theory of existence and uniqueness of solutions to a time–dependent DE’s of the general form in \((0, T) \times \mathbb{R}^d, d \geq 1 \) is an integer, refers to the study of \( \dot{y}(t) = v(t, y(t)) \)
under prescribed \( y(0) = x \). It gives sufficient conditions on the vector field \( v(t, x) : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d \) that guarantee existence (and, maybe, uniqueness) of solutions for some initial data \( x \in \mathbb{R}^d \) and some (small) time. More precisely, the classical theory of flows corresponding to maps \( v \), which are Lipschitz continuous in space uniformly with respect to time, has become known as the Cauchy–Lipschitz theory. It basically shows (Picard–Lindelöf Theorem) that, the solution \( X(t, \cdot) \) of the initial value (or Cauchy) problem

\[
\partial_t X(t, \cdot) = v(t, X(t, \cdot)), \quad X(0, \cdot) = \text{Id}
\]

exists and it is unique for all times \( t \in [0, T] \). For mere existence the continuity of \( v \) is sufficient (Peano’s Theorem). Uniqueness is also obtained under less stringent conditions such as one–sided Lipschitz condition or Osgood continuity [41]. If \( v \) is sufficiently regular, then \( X(t, \cdot) \) is a diffeomorphism of \( \mathbb{R}^d \). More generally, if \( v \in C^1_c([0, T] \times \Omega) \), where \( \Omega \subset \mathbb{R}^d \) is open, then the unique flow map \( X(t, \cdot) \) is a diffeomorphism of \( \Omega \) [95], [113].

Note that the solution \( X \) is written as a function of two variables in order to account not only for the time–variable but also for the initial value prescribed for \( X \). In the case where the solution exists and is unique for all \( x \in \mathbb{R}^d \), loosely speaking, the function \( X \) is called the classical flow of \( v \). The terminology comes primarily from Fluid Dynamics: if \( v \) stands for the velocity of fluid flow, then \( X(t, x) \) accounts for the position at time \( t \) of the fluid particle that was initially \( (t = 0) \) at position \( x \). This equation is the basis for the Lagrangian description of fluid flow (where the trajectory of a particle is observed through time). Thus, we take the opportunity to further motivate our study by pointing out that it addresses the fundamentals of Lagrangian description of Fluid Dynamics.

The flow equation above is closely related to the continuity equation from Fluid Dynamics

\[
\partial_t \mu(t, \cdot) + \nabla \cdot (\mu(t, \cdot) v(t, \cdot)) = 0,
\]

coupled with the initial data \( \mu(0, \cdot) = \mu_0 \), where the curve \( (0, T) \ni t \to \mu(t, \cdot) =: \mu_t \) is a family of positive locally finite (Borel) signed measures on \( \mathbb{R}^d \), which depends on the time parameter \( t \in (0, T) \). This equation is intended in distributional sense and \( \mu_t \) is called a measure–valued (or, a weak) solution if

\[
\int_0^T \int_{\mathbb{R}^d} \left( \partial_t \xi(t, x) + v(t, x) \cdot \nabla \xi(t, x) \right) d\mu_t(x) dt = -\int_{\mathbb{R}^d} \xi(0, x) d\mu_0(x).
\]
for every bounded and Lipschitz test function $\xi \in C^1_c([0,T] \times \mathbb{R}^d)$ (see Definition 1.7). Indeed, in a sufficiently smooth setting (see Theorem 1.18), the measures $\mu_t$ are given by

$$
\mu_t = X(t, \cdot) \# \mu_0, \quad \text{i.e.} \quad \int_{\mathbb{R}^d} \xi \, d\mu_t = \int_{\mathbb{R}^d} \xi \circ X(t, \cdot) \, d\mu_0$$

for all $t \in [0,T]$, $\xi \in C_c(\mathbb{R}^d)$, where $X(t, \cdot) \# \mu_0$ denotes the pushforward of the measure $\mu_0$ via the map $X(t, \cdot) : \mathbb{R}^d \to \mathbb{R}^d$; that is, $\mu_t(B) = \mu_0 \circ X^{-1}(t, \cdot)(B)$ for any Borel set $B \subset \mathbb{R}^d$, according to Definition 3.2. If the identity $\text{Id}$ is replaced by an arbitrary map $X_0$, all the observations mentioned above remain valid in the general, multidimensional case, the only different is that $\mu_0$ will be replaced by $X_0 \# \mu_0$ as the initial point on the curve $(0,T) \ni t \to \mu_t$.

The smooth setting becomes, however, unsatisfactory when one is looking at problems from mathematical physics, mainly systems of conservation laws (we recommend [4], [6], [14], [41] for a concise outlook on this matter). The starting point is that a system of conservation laws can be thought of as a transport system in which $v$ will depend on the actual density $\rho$. Due to formation of shocks (characteristics crossing), there is no “smooth” theory when it comes to linking the flow equation to the transport equation

$$
\partial_t F(t, \cdot) + v(t, \cdot) \cdot \nabla F(t, \cdot) = 0, \quad F(0, \cdot) = F_0.
$$

To prove existence of essentially bounded solutions for the flow problem is easily achieved by smooth approximation even in the case of essentially bounded $v$ with locally integrable distributional divergence. In the Sobolev case, DiPerna and Lions published a celebrated paper [56] in which they characterized results which relates renormalization, strong continuity, uniqueness and approximation with smooth functions. It basically means that, whenever $F$ is a weak solution for the transport equation, then so is $\phi(F)$ for smooth $\phi$. Obviously, classical solutions, if they existed, would satisfy that. Ambrosio [6] took the question one step further and developed a uniqueness theory based on the renormalization property of $v$, but only assumed spatial bounded variation regularity for the vector field. More precisely, DiPerna–Lions [56], Ambrosio [6] (also, see [41] for a good survey on such problems) addressed the questions of existence, uniqueness and stability for regular Lagrangian flows, that is, solutions $X$ of the flow equation such that $X(t, \cdot) \# \mathcal{L}^d \ll C\mathcal{L}^d$ for some positive real constant independent of $t \in [0,T]$. In the Sobolev case (there is a similar version if only certain $BV$ regularity on $v$ is assumed), the almost (minor improvements are available [41])
state–of–the–art uniqueness result covers only velocities

\[ v \in L^\infty((0, T) \times \mathbb{R}^d; \mathbb{R}^d) \cap L^1([0, T]; W^{1,p}(\mathbb{R}^d; \mathbb{R}^d)) \] for some \( p > 1 \).

It is proved that if \( v \) is this regular, then its regular Lagrangian flow, if it exists, then it is unique. Thus, their study is an extension of the classical theory of existence and uniqueness of solutions for differential equations. It is, therefore, different from ours, as we will not be imposing conditions on the velocity, but rather on the Eulerian flow whose Lagrangian characterization is provided by the flow equation. Note that, the boundedness assumption on the vector field assumed above is not essential: in this work we will also address the theory of Lagrangian flows relative to vector fields satisfying more general (Borel) regularities. We had originally planned a complete departure from that setting, as our initial goal will be to investigate the uniqueness of the Lagrangian flow associated to a given absolutely continuous path of probability measures (see Definitions 5.1 and 6.2 below). The main purpose of this work is to present a similar result to the above but within a different approach and by using new techniques.

The problem we address here comes from the different perspective. Our study will be completely different than those of [6], [56], as the main topic of this thesis is to understand if solutions exist and/or they are unique as functions of both time and initial data. In what follows, we shall describe our current results and what else we believe can be further achieved. Let us begin by investigating the DE

\[ \partial_t X(t, x) = v(t, X(t, x)), \quad \text{under } X(0, x) = X_0(x), \quad x \in I, \]  

where \( I \) is the interval \((0, 1)\) and \( X_0 : I \to \mathbb{R}, v : (0, T) \times \mathbb{R} \to \mathbb{R} \) are given functions. As mentioned earlier, this equation describes the flow of a scalar field when the vector field \( v \) represents the fluid flow velocity at position \( x \) (if \( X_0 \equiv \text{Id} \)) or, more generally, \( X_0(x) \) for every \( x \in I \), at time \( t \), and the fluid flow is described by a map \( X_t := X(t, \cdot) \) giving the position of the particle labeled \( X_0 \) at time \( t \). We ask the question of whether given a Borel velocity field \( v : (0, T) \times \mathbb{R} \to \mathbb{R} \), there exists a (unique) family of flow maps \( X : (0, T) \times I \to \mathbb{R} \) whose velocity is \( v \). Of course, \( X \) will have to be weakly differentiable in time for this question to even make sense. We are thus looking for characteristic (or integral) curves of the given velocity field \( v \); that is, curves with the property that at each point the tangent
vector coincides with the value of the given vector field at such point.

In this work we consider the case of $\chi$ being the one–dimensional Lebesgue measure restricted to the unit interval of $\mathbb{R}$. Note that, $\chi$ is a Borel probability measure on $\mathbb{R}$ and $\mu_t := X_t\#\chi$ gives the Eulerian description of the uniform distribution $X_0\#\chi$ on $I$ as it is transported by the velocity $v$: this means $(\mu, v)$ solves the unidimensional continuity equation

$$\partial_t \mu + \partial_x (\mu v) = 0 \quad \text{in} \quad (0, T) \times \mathbb{R}$$

from Fluid Mechanics in the sense of distributions. It is worth mentioning that (Flow) and (CE) are closely connected to the unidimensional transport equation

$$\partial_t F + v \partial_x F = 0 \quad \text{in} \quad (0, T) \times \mathbb{R}$$

even though we shall not pursue this in our present study. Indeed, the cumulative distribution function $F_t$ of the density $\rho_t$ formally satisfies the above equation (see Proposition 7.3 below).

The motivation for our problem arose recently in joint work of the author with A. Tudorascu [21]. It concerns the well–posedness theory of the partial differential equations of the Lagrangian and Eulerian problems type. More precisely, [21] studies existence, uniqueness and regularity of solutions for (Flow) and (CE) within a certain class of curves associated to a given velocity. The velocity $v$ in the continuity equation above is related to the optimal maps pushing $\chi$ forward to the current measure $\mu_t$ on the solution curve that is, indeed, absolutely continuous (see Proposition 5.8 below). The flow–map formula (5.10) relating these optimal maps and $v$ is essential to proving that the weak solutions for (Flow) constructed in [21] are unique. In [21], the same formula is the main ingredient for the argument that a time–dependent rearrangement of monotone solution for (Flow) obtained by Brenier [34] does, indeed, satisfy (Flow).

Let us illustrate some basic facts from the theory of differential equations in the smooth framework. If $v = v(t, y)$ is sufficiently smooth, one can formally take the derivative of both sides of (Flow) with respect to $x$ to verify that the spatial derivative $u = \partial_x X(t, x)$ satisfies

$$\partial_t u(t, x) = \partial_y v(t, X(t, x)) u(t, x), \quad \text{under} \quad u(0, x) = X'_0(x), \quad x \in I.$$
The solution procedure shows that

$$u(t, x) = X'_0(x) \exp \left( \int_0^t \partial_y v(s, X(s, x)) \, ds \right).$$

Whereas we have not really found an explicit solution (because $\partial_y v$ depends on $X$), we can conclude that $X'_0(x) > 0$ implies $\partial_x X(t, x) > 0$ for all $(t, x) \in (0, T) \times I$, which shows that for every $t \in (0, T)$, the function $x \to X(t, x)$ is strictly increasing, with positive slope everywhere. This, when referring to (CE), is in agreement with the intuition that “nice” velocities $v$ preserve in time the order of the positions of particles on the real line, or, equivalently, the characteristics do not cross.

If $v$ is not smooth enough to justify (Diff) and our little calculation above, then the order need not be preserved through time. This has fundamental implications in scalar Conservation Laws, where the crossing of characteristics is responsible for the formation of shocks. Our goal in this thesis is to study conditions under which $x \to X(t, x)$ stays monotone nondecreasing for all times $t \in [0, T]$ if $X_0$ is itself monotone nondecreasing. The meaning of “conditions” is quite vague at this point; however, they shall not be imposed on the velocity $v$ itself, but rather on the absolutely continuous curve of probabilities (CE) whose velocity $v$ is given by (Flow). Under the minimal assumptions we shall impose on $\mu$, the velocity $v$ will be uniquely determined by $\mu$ from (CE) (in some precise sense). Thus, our endeavor will be to study (Flow) under the constraint that $X_t \# \chi = \mu_t$ be given for all $t \in [0, T]$ in the form of an absolutely continuous curve of probability measures [17].

Motivated by the well–known results of [6], [56] that for a given velocity vector field, which is merely a Borel function, we define flows of Borel maps, that is, the absolutely continuous maps $[0, T] \ni t \to X(t, x)$ that satisfying (Flow) for a.e. $x \in I$ (see Definition 5.1 below). The reason for introduce this notion will become clear later in Chapter 5 since it turns out to be the right one in the study of (Flow) with weakly differentiable solutions. In fact, this definition is different from the one described in [6], [56] (other than the fact that we restrict our study to the one–dimensional case), we only consider the initial spatial positions to lie in a bounded interval. Note that, we take $I := (0, 1)$ to ensure that the measure $X \# \chi$ (where $\chi := L^1|_I$) is a probability measure, but everything we achieve in this thesis can be trivially extended to any bounded interval.

As mentioned above, our focus will be different of [6], [56] as we will not be looking for conditions on $v$ that render uniqueness (within a certain class, such as regular flows, as
described above). Instead, we would like to characterize solutions of \( \text{(CE)} \) which admit a unique Lagrangian description. Thus, no explicit conditions on the velocity \( v \) were to be made, even though any assumption on the curve of measures will implicitly reflect on the (unique) velocity associated to it. We have managed to stay true to this goal only in part; see the main result Theorem 7.14, which covers the case of higher integrability for the densities on the curve. However, while analyzing the case of continuous densities (Theorem 7.5, the other main result) we discovered that a continuous velocity would not only yield uniqueness for the Lagrangian flow associated to the curve, but it would also give uniqueness for \( \text{(CE)} \) within the class of absolutely continuous curves of probabilities [17]. Thus, in Theorem 7.5 we also make the continuity assumption on \( v \).

We would like to make it clear that we deal only with solutions of \( \text{(CE)} \) that are absolutely continuous curves of probability measures: it is easy to construct examples of velocities \( v \) and distributional solutions of \( \text{(CE)} \) originating at a given probability \( \rho_0 \), but whose mass changes in time, and/or become negative. Indeed, one may take any smooth and positive density \( \rho_0 \) that vanishes at \( \pm \infty \), and set \( \rho(t, y) := (1-t)|1-t|\rho_0(y) \): then \( \rho \) is a distributional solution for \( \text{(CE)} \) with \( v(t, y) = 2F_0(y)/((1-t)\rho_0(y)) \) if \( t \in [0, 1) \cup (1, \infty) \) and \( v(t, \cdot) \equiv 0 \) if \( t = 1 \) (\( F_0 \) is the cumulative distribution function of \( \rho_0 \)). Note that while \( \rho(0, \cdot) \) is a smooth probability density, the mass of \( \rho(t, \cdot) \) decreases in time, and \( \rho(t, \cdot) \) even goes below zero for \( t > 1 \). Such solutions of \( \text{(CE)} \) are non–physical in Conservation Laws or, for example, Thermodynamics, where this equation is also used to express precisely conservation of mass for \( \rho \); furthermore, as \( \rho \) usually denotes a physical quantity (such as material density or absolute temperature), it is also important in applications that it does not pointwise below zero.

Let us recall some basic facts from the theory of \( L^p \)-absolutely continuous curves in \( \mathcal{P}_p(\mathbb{R}^d), 1 \leq p < +\infty \) (the space of Borel probability measures on \( \mathbb{R}^d \) with finite \( p \)-th–order moments). We shall be quite sketchy, but we shall review what little there is in Chapter 4, and for further details we recommend the comprehensive reference [17]. Let us endow \( \mathcal{P}_p(\mathbb{R}^d) \) with the \( p \)-th–Wasserstein metric defined by

\[
W_p^p(\mu, \nu) := \min_{\gamma} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, d\gamma(x, y),
\]

where the infimum is taken among all probabilities on the product space \( \mathbb{R}^{2d} \) with marginals \( \mu, \nu \). Thus, \( (\mathcal{P}_p(\mathbb{R}^d), W_p) \) becomes a \textit{Polish space} on which we define absolutely continuous...
curves by saying that \([0, T] \ni t \to \mu_t := X_{t \#} \chi \in \mathcal{P}_p(\mathbb{R}^d)\) lies in \(AC^q(0, T; \mathcal{P}_p(\mathbb{R}^d))\), \(1 \leq q \leq +\infty\), provided that there exists \(\beta \in L^q(0, T)\) such that \(W_p(\mu_t, \mu_{t+h}) \leq \int_t^{t+h} \beta(s) \, ds\) for all \(0 < t < t+h < T\). The metric derivative of such a curve is defined as

\[
|\mu'(t)| = \lim_{{s \to t}} \frac{W_p(\mu_s, \mu_t)}{|s-t|} \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, T).
\]

It is proved in the book [17] of Ambrosio, Gigli and Savaré that, if \(1 < p < +\infty\), then any such curve admits a unique Borel velocity field \(v : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d\) such that \((\mu, v)\) satisfies (CE) in the sense of distributions in \((0, T) \times \mathbb{R}^d\), and \(v(t, \cdot) \in L^p(\mu(t, \cdot); \mathbb{R}^d)\) for a.e. \(t \in (0, T)\). The most general problem (Flow) discussed here in this thesis assumes only Borel regularity on \(v\) and we only study solutions \(X\) belonging to some time–Sobolev spaces \(W^{1,q}(0, T; L^p(I))\) for \(1 \leq q \leq +\infty\) and \(1 \leq p < +\infty\). The reason for this extra–requirement on the object defined in Definition 5.1 will become clear in Chapter 5, where we prove that any map \(X\) as in Definition 5.1 which also satisfies \(X_{t \#} \chi = \mu_t\) for all \(t \in [0, T]\) for some \(\mu \in AC^q(0, T; \mathcal{P}_p(\mathbb{R}))\) must, in fact, lie in \(W^{1,q}(0, T; L^p(I))\). This also means that the time–derivative along paths \(X(\cdot, x) \in W^{1,1}(0, T)\) which is denoted by \(\partial_t X(\cdot, x) \in L^1(0, T)\) coincides \(\mathcal{L}^2\text{–a.e.}\) with the functional derivative \(X \in L^q(0, T; L^p(I))\) of \(X\). This definition turns out to be equivalent to requiring that \(X \in W^{1,q}(0, T; L^p(I))\) satisfy the integral equation

\[
X(t, x) = X_0(x) + \int_0^t v(s, X(s, x)) \, ds \quad \text{for } \mathcal{L}^1\text{-a.e } x \in I \text{ and every } t \in [0, T].
\]
It was proved in [70] that, if \( v \) is the velocity associated to some curve \( \mu \in AC^2(0, T; P_2(\mathbb{R})) \) and \( M_t : I \to \mathbb{R} \) is a monotone nondecreasing map pushing \( \chi \) forward to \( \mu_t \) for every \( t \in [0, T] \), then the family of optimal maps \( M_t, t \in [0, T] \), say \( M \), belongs to the Hilbert space \( H^1(0, T; L^2(I)) =: \mathcal{H} \) and satisfies (Flow) in the sense of Definition 5.1. The proof [70] is based on the observation that for Lebesgue almost all \((t, x), (t, y) \in (0, T) \times I, M_t(x) = M_t(y) \) implies \( \dot{M}_t(x) = \dot{M}_t(y) \) (in case both derivatives exist pointwise in some sense, later to be specified). Furthermore, the proof uses the fact [70] that the metric derivative \( |M'(t)| \) exists at \( t \in (0, T) \) if and only if the metric derivative \( |\mu'(t)| \) exists at \( t \); in that case \( |M'(t)| = |\mu'(t)| \).

Note that, maps in \( \mathcal{H} \) possess a “functional” time–derivative in \( L^2((0, T) \times I) \), which, in fact, coincides on a subset of \((0, T) \times I \) of full Lebesgue measure with a pointwise “pseudo”–derivative [70]. Equation (Flow) is to be understood in that context; that is, it should be satisfied everywhere in the set where the functional derivative \( \dot{X} \) coincides with the pointwise pseudo–derivative. Of course, this definition also turns out to be equivalent to requiring that \( X(\cdot, x) \) satisfy the integral identity above. Moreover, it has recently been proved [21] that, if \( X_0 \) is monotone nondecreasing and a map \( X \in \mathcal{H} \) solves (Flow), then so does \( \text{Mon} \), where \( \text{Mon}(S) \) stands for the monotone rearrangement of the map \( S \). The proof of this result is also based on tools developed in [70], [92]. Notwithstanding the imposed \( \mathcal{H} \)–regularity, the previous statement basically states that whenever (Flow) has solution \( X \), then it will automatically have a monotone solution \( M \) (which, in this context, we take it to mean that \( M_t \) is monotone nondecreasing for all \( t \in [0, T] \)) obtained by setting \( M_t := \text{Mon}(X_t) \) for all \( t \in [0, T] \).

The concept of “monotone rearrangement” makes the connection with the theory of one–dimensional Optimal Transport. It has been known since Fréchet [67] that, any Borel map \( S \) defined on (in our case) \( I \) can be monotonically rearranged over \( I \); that is, there exists a nondecreasing map \( M : I \to \mathbb{R} \) such that \( \chi(S^{-1}(B)) = \chi(M^{-1}(B)) \) for all Borel sets \( B \subset \mathbb{R} \). In other words, there exists a nondecreasing function \( M \) such that the Lebesgue measure restricted to \( I \) of the preimages of any Borel set through \( S \) and \( M \) coincide. By one–dimensional Optimal Transport, it is known that there is a bijective correspondence between the closed, convex cone of monotone nondecreasing functions in \( L^2(I) \) and the metric space \( P_2(\mathbb{R}) \) endowed with the quadratic Wasserstein distance (see, e.g., [70]). This implies [70] (due to the uniqueness of the velocity for a given \( AC^2 \) curve) there is a bijective correspondence between paths \( M \in \mathcal{H} \) consisting of monotone nondecreasing maps \( M_t, t \in [0, T] \) and curves \( \mu \in AC^2(0, T; P_2(\mathbb{R})) \), via \( M_t \# \chi = \mu_t \). Thus, we can now combine this fact
with the knowledge that each solution to (Flow) gives rise to a monotone solution to state the important consequence: The initial value problem (Flow) admits a solution in $\mathcal{H}$ if and only if $v$ is the velocity associated with some curve $\mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R}))$. This may look like a very strong statement, yet its practical importance from the point of view of existence for (Flow) relies on the fact that one knows a priori that a Borel function $v : (0, T) \times \mathbb{R} \to \mathbb{R}$ satisfies (CE) for $\mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R}))$ and $v \in L^2(\mu_t)$ for a.e. $t \in (0, T)$. We shall see in Chapter 5 that one can obtain such a curve from (Flow) by defining $\mu_t := X_t \# \chi \in AC^q(0, T; \mathcal{P}_p(\mathbb{R}))$; thus, we shall see that if $X_0$ is nondecreasing, then as soon as (Flow) admits a solution $X \in W^{1,q}(0, T; L^p(I))$, it automatically admits the solution $M \in W^{1,q}(0, T; L^p(I))$ consisting of the monotone rearrangements of the maps $X_t$.

A study of the regularity of the monotone rearrangements of maps belonging to a time-continuous family of maps was performed by Loeper in [85]. It is proved that the distributional time derivative is a signed measure. This is very weak by comparison to what we achieve here, but it covers a much more general case: arbitrary spatial dimension, and the family of maps is not necessarily the flow of a map.

Let us now return with a comment back to (CE). Note that, (CE) will still be satisfied by $\mu_t := X_t \# \chi$ in the sense of distributions if $X \in \mathcal{H}$ is such that $X_t$ is $\chi$-essentially injective for $\mathcal{L}^1$-a.e. $t \in (0, T)$. The velocity can be given by $v_t := \dot{X}_t \circ X_t^{-1}$, a well-defined map on the support of $\mu_t$ (which we assume to be Borel measurable in time–space). The maps $M_t$ are monotone nondecreasing rearrangements of $X$ and for Lebesgue almost all $(t, y) \in (0, T) \times \mathbb{R}$ the fibers $X_t^{-1}(y)$ are either singletons or have positive one-dimensional Lebesgue measure. Thus, the essential injectivity assumption on $X$ can be relaxed as in [114].

We have seen earlier that any solution $X$ of (Flow) will be nondecreasing in $x$ provided that $X_0$ is monotone nondecreasing, $X$ and $v$ are regular enough to justify (Diff). It is, however, this regularity that might not be there, which will enable solutions that are not nondecreasing in $x$ (see a first example in Chapter 5, Example 5.10). The other examples, Examples 5.13 and 5.14 however, illustrate that there are even irregular $v$ for which only nondecreasing solutions exist. Thus, it comes as a very natural question to investigate when (Flow) has only spatially nondecreasing solutions. The subtlety of the matter will become very clear, as we shall see that a careful formulation is in order.

Notice that, Example 5.10 below is an example where (Flow) has other solutions, beside monotone nondecreasing ones. It also illustrates that once a solution exists, there will also exist a monotone nondecreasing solution, as discussed above. The obvious issue with that
example is that, whereas for $t > 1$ we get decreasing solutions, the solution $X_t$ is still nondecreasing (and, thus, coincides with $M_t$) for all $t \in [0, 1]$. It is not difficult to modify the example in such a way that the time threshold $t = 1$ is replaced by an arbitrarily small positive time. However, the natural question is whether it is true that for any solution $X \in W^{1,q}(0, T; L^p(I))$ of (Flow) with $X_0 \equiv \text{Id}$ there exists a time horizon $\bar{t} > 0$ such that $X_t$ is nondecreasing for all $t \in [0, \bar{t}]$. As Example 5.12 will show, this is false. Following a suggestion by Alberto Bressan, we have constructed a velocity $v$ and its flow $X$ such that for any $\varepsilon > 0$ there exists a time $0 < t_\varepsilon < \varepsilon$ for which $X_{t_\varepsilon}$ is not nondecreasing. Note also that, not only does Example 5.12 provide an instance where the nonuniqueness of the flow (albeit classical) manifests instantaneously for $t > 0$ (as we have seen there are solutions other than the monotone ones for arbitrarily small positive time), but it also gives instantaneous mass concentrations (Dirac deltas) in the measures $X_{t\#}\chi$ (as a result of $X_t$ developing flat portions instantaneously). Also, Example 5.13 shows that this can occur even if the flow is unique (i.e. consisting of monotone nondecreasing maps). More precisely, $X_t$ may develop “flat portions” as soon as $t$ becomes positive; that is, $X_t$ does not have to remain strictly increasing for small times. Furthermore, as Example 5.14 shows, it is also possible that there are two different monotone solutions $X$ and $Y$ for (Flow). The example is constructed along the lines of the classical $v(t, y) = \sqrt{|y|}$, which is often used to illustrate nonuniqueness for the initial value problem $\dot{y} = \sqrt{|y|}$, $y(0) = 0$. This example helps justify why the constraint “$X_{t\#}\chi$ is prescribed” is used here as a criterion to sort out between solutions. In [6], the author uses this example towards the same goal. The criterion chosen there is different from ours: it is the regularity of the flow, as described earlier.

As far as uniqueness for (Flow) goes, Example 5.14 shows that in general there might be more than one curve whose velocity $v$ is. This accounts for a nonuniqueness “mechanism”, which we now dub multiple–curve (MC) condition on the velocity $v$. Thus, Example 5.14 provides a $v$ satisfying (MC) (or, we could say “$v$ is (MC)”). If $v$ is not (MC), then we say $v$ is (SC) (single–curve). Examples of $v$ with (SC) are abundant; any continuous $v$ that is also Lipschitz in the $x$–variable would do, as this implies pointwise existence and uniqueness for the ordinary differential equation $\dot{y}(t) = v(t, y(t))$ for any initial data $y = y(0)$. But this regularity is not necessary, as we have seen in Remark 5.18 that the velocity from Example 5.13 produces a single flow map $X$, since it is (SC) (even though it is discontinuous) and its (unique) curve develops atoms as soon as $t > 0$. Thus, if a velocity $v$ is (MC), there is no hope for a uniqueness result for (Flow), as each curve of probabilities associated with $v$ will
produce its own distinct solution (consisting of the optimal maps). The next question is: since $v$ being (MC) leaves no hope for a uniqueness result for (Flow), how about uniqueness of solutions $X$ for (Flow) under the constraint $X_t \# \chi = \mu_t$ for $t \in [0, T]$ (where $\mu$ is a curve transported by $v$ via (CE))? By the statement discussed above and Proposition 4.2 [70] we know the optimal maps $M_t$ such that $M_t \# \chi = \mu_t$ give a solution. We are naturally led to:

Under what conditions on $\mu$ do we have that the path $M \in W^{1,q}(0,T;L^p(I))$ consisting of the optimal maps $M_t$ such that $M_t \# \chi = \mu_t$ is the unique solution to (Flow) which pushes $\chi$ forward to $\mu$, where $v$ is the $L^p(\mu)$ velocity map along the curve $\mu \in AC^q(0,T;\mathcal{P}_p(\mathbb{R}))$ for some $T > 0$ and such that $\mu_0 = \chi$? This question of whether the family of optimal maps is the only solution becomes pertinent again once we prescribe the curve associated with $v$. This is yet another motivation for undertaking the present analysis.

Our main contribution is identifying sufficient conditions on a path $\mu \in AC^q(0,T;\mathcal{P}_p(\mathbb{R}))$ to ensure that the path $M \in W^{1,q}(0,T;L^p(I))$ consisting of the optimal maps such that $M_t \# \chi = \mu_t$ is the unique solution to (Flow) which pushes $\chi$ forward to $\mu$. As Example 5.10 refers to an (SC) velocity, the nonuniqueness in that case is solely due to having other solutions (other than the one consisting of optimal maps) corresponding to the unique curve of measures associated to the velocity $v$. As we can easily compute the relation (5.12), the curve $\mu$ consists of the singular measure (Dirac mass) in the middle. In terms of the maps $X_t$, we see that $X_t$ becomes flat before becoming not nondecreasing (i.e. at time $t = 1$). We have not been able to construct an example where $X_t$ changes behavior from nondecreasing to not nondecreasing without becoming flat first (which translates into $\mu_t$ having atoms). To exclude the possibility of existence the singular measure in spatial dimension $d = 1$, we have addressed the questions of existences and/or uniqueness for Lagrangian flows of a Borel maps $X$ which satisfies (Flow) in the sense of Definition 5.1 (see Theorems 7.5 and 7.14). In the Hilbert space $\mathcal{H}$, see Theorem 5.20, we have proved a much stronger result than in [6], [41], [56] for (SC) velocities. The proof of these results relies on the polar decomposition [34] of maps $X$ such that $X_t \# \chi = \mu_t$, i.e. maps $g_t$ such that $X_t = M_t \circ g_t$ and $g_t \# \chi = \chi$. As a byproduct of our approach, uniqueness for the continuity equation (CE) also arises within a precisely defined class of solutions; within that class, the velocity $v$ will be an (SC) velocity.

Note, however, that there is a trade–off between regularity assumptions on $v$ and specifying the pushforward measure. We have been able to prove that the measure–preserving map $g_t$ is; indeed, a time–independent and equal to the identity map of $I$. We have discovered that the requirements on $\mu$ can, in fact, be relaxed, but it is not clear to us yet to what
extent. It seems likely that the requirement that $\mu_t$ have no atoms for all $t$ is sufficient. It is, therefore, natural to ask whether a nonatomic $\mu_t$, for all $t \in [0, T]$, guarantees the uniqueness we are looking for. In other words: Is it enough to require that $\mu_t$ have no atoms for the monotone solution to (Flow) to be the only solution? Note that, Example (5.13) clearly shows, however, that $\mu$ being atomless is not necessary for uniqueness.

Lastly, it is not clear whether the constraint on $X_t$ pushing $\chi$ forward to a specified $\mu_t$ enjoying certain regularity is necessary. We do not have an example of a velocity shared by two different curves that do not contain atomic measures. The nonatomic property of $\mu$ is equivalent to $M_t$ having no flat portions (i.e. no intervals where it is constant), or the cumulative distribution function of $\mu_t$ being continuous.

We expect that our study will have a strong impact on how first–order, scalar differential equations will be viewed and analyzed in the future. As we have already pointed out, the idea of treating these DE’s as depending on the parameter given by the initial data is not new [6], [41], [56]. This should come as no surprise, given that such equations describe the flow of a scalar field, which is the reason why they are instrumental to the Lagrangian description of particle motion. However, the proposed approach is novel in that it is based on the connection between the flow equation and one–dimensional Optimal Transport. We have seen that new results (requiring less regularity on the velocity map) are emerging this way, and we expect to discover more. In our opinion, exploring the close link between (Flow), (CE), and (Trans) is worth pursuing before more complicated, higher–order PDE’s are studied. The main idea here stems precisely from the analysis of ODE’s of the type (Flow) performed by taking into account the dependence of solutions on initial data.

If the velocity $v$ is a “nice” function, the whole theory is simple and well–understood. Otherwise, the misleading simplicity of (CE) does not warn the researcher of the hazards involved with its analysis. Its connection with (Flow) offers a glimpse into that, since (Flow) can be too general to be handled easily. These issues have been recognized and have been the object of deep mathematical works [6], [41] etc. Our approach in the present work has the potential to fill some knowledge gap even in the simplest of cases (spatial dimension one).

We finally illustrate the content of the various chapters of this thesis. For more details the reader is referred to the introduction of each chapter and each section.

Chapter 1 presents Lagrangian and Eulerian descriptions of fluid flow: we start with the Lagrangian and Eulerian points of view and how to describe a transportation phenomenon via
the flow. Then we introduce a related partial differential equation, the continuity equation, and for it we study measure–valued solutions; we present, in the smooth case, the connection between the continuity equation and the ODE, which is based on the classical theory of flows. Subsequently, starting from Section 1.4 we begin to consider the nonsmooth case from the Eulerian viewpoint. We see first that existence of weak solution holds under very general hypotheses. The notation of renormalized solutions is then introduced, together with its consequences for the well–posedness of the PDE, by DiPerna and Lions [56] in the Sobolev case and by Ambrosio [6] in the $BV$ case. Then we comment on other various renormalization results and on the case of nearly incompressible vector fields.

Chapter 2 is devoted to the theory of regular Lagrangian flows that describes the connection between the Lagrangian problem (ODE) and Eulerian problem (PDE) out of the smooth setting. First subsection is devoted to the notion of probabilistic solution and to the discussion of the probabilistic representation, which extends the theory of characteristics to the nonsmooth case. Then, the equivalence between pointwise uniqueness for the ODE and uniqueness of positive measure–valued solutions of the continuity equation is shown, exploiting the probabilistic representation for solutions of the PDE. We then introduce the concept of regular Lagrangian flow; the “good” notion of solution to the ODE in the nonsmooth context, and we show how the uniqueness of bounded solutions of the continuity equation imply the well–posedness of the regular Lagrangian flow. We finally comment briefly on the DiPerna–Lions’ notion of flow and on the notation of density of a regular Lagrangian flow.

Chapter 3 presents the relaxation that Kantorovich applied to the original Monge problem and its duality issues (Kantorovich potentials, $c$–cyclical monotonicity, etc.). It uses these tools to provide the first theorem of existence of an optimal map (Brenier theorem). The purpose of this chapter is to familiarize the reader to some extent with the Monge–Kantorovich theory of optimal transportation. Since the Wasserstein distance (metric) and optimal transportation maps constitute the main tools in our research, we feel that the underlying theory is worth reviewing in more or less detail. The Monge problem and Kantorovich’s relaxed version, along with Kantorovich’s duality principle are the motivations of what is developed next. So is Brenier’s theorem, perhaps the most important step undertaken towards the resolution of the Monge problem. These results are stated without proofs and are accompanied by a relevant (for our purpose) discussion and appropriate references.

The unidimensional case of optimal transport problem, which is easier and already has many consequences, is analyzed in detail in Chapter 4, in which we mainly follow [104]. The
generalized inverse of the cumulative distribution function of the measures on the curve is presented; it is a monotone transport map, and its optimality is discussed. Then we show that the Monge–Kantorovich transport problem can be solved with various types of the cost function. We show existence and uniqueness results of the optimality of monotone transport maps and plans “line by line”, using explicit arguments and a careful study of the structure of the problem and its costs functions in the 1D case, and this eventually implies uniqueness for the optimal transport maps. Additionally, Monge–Kantorovich metrics on the real line are discussed. There are different ways of defining the Wasserstein distance. One is, naturally, introduced as the optimal totals cost from the Monge–Kantorovich theory. We have chosen, due to applying 1D optimal transport theory, to introduce the Wasserstein distance as the $L^p$–norm of the monotone maps. An entire section is devoted to its properties and more or less detailed proofs are given.

With Chapters 5 and 6 we start to present a different look at the Lagrangian viewpoint. These chapters begin by briefly reviewing and recalling definitions of some objects and essential concepts to our presentation, such as Wasserstein distance/space, absolutely continuous curves of probability measures, generalized inverse etc. Some nontrivial and interesting examples which verify uniqueness and nonuniqueness of flows for several vector fields are provided in Chapter 5, including a counterexample showing a vector field for which there exists a nonmonotone flow on any subinterval. Then a necessary and sufficient condition on the regularity Lagrangian flow is proved in order for the associated path of measures to possess $AC$-regularity. The main results of Chapter 6 are Theorems 6.33 and 6.35 which show that quite generally $AC$ paths admit Lagrangian descriptions provided by the family of optimal maps.

Chapter 7 focuses on two different types of conditions under which the said Lagrangian description is unique: Theorem 7.5 explores the case of continuous densities and velocities by a direct method (loosely connected to the narrative from Chapter 5), whereas Theorem 7.14 is more deeply indebted to the results from Chapter 5 as it analyzes the case where the densities enjoy some precisely quantified integrability (no conditions imposed on the velocity in this case). Theorem 7.5 also provides a uniqueness result for the continuity equation (CE) within a reasonably general class of solutions. Theorem 7.14 does the same, albeit in a more restrictive setting. In both cases, the uniqueness results for the Lagrangian flow and the uniqueness results for the continuity equation are obtained concomitantly from our method of proof.
Chapter 1

An overview on Lagrangian and Eulerian descriptions of fluid flows

In this chapter, we are going to describe the connection between Lagrangian and Eulerian descriptions of a fluid flow. We do not try to give here an account of the extremely wide literature on this topic, but we rather prefer to illustrate with some detail a specific case which is presented in [6], [56] and which was a motivation for the main results collected in this thesis, namely DiPerna–Lions and Ambrosio’s notions (see Theorems 1.22 and 1.31). In Section 1.1, we present a very informal introduction to the physical interpretation of Lagrangian flows in Fluid Dynamics. In Sections 1.2 and 1.3, we present some ideas and the main results of the classical theory of Lagrangian flows, which relates the two problems of the Eulerian problem (PDE), namely the continuity and the transport equations, and the Lagrangian problem (ODE) in the case of a sufficiently smooth vector field: we summarize and recall the most important results of the theory of ODE in the smooth framework where a concept of the classical flow can be given to study existence, uniqueness and stability of solutions. Most of these results are very classical, hence we give only a sketch of the proof. Basic references for these topics are for instance [12], [17], [104], [31], [74], [111]. In Section 1.4, we begin to investigate the well–posedness of the the continuity equation out of the smooth setting. After defined rigorously, in Subsection 1.3.1, the notion of weak (distributional) and measure–valued solution, which is needed in order to consider the nonsmooth case, it is shown, in Subsection 1.4.1, a very general result of existence of bounded solutions.
Subsections 1.4.2 and 1.4.3 are devoted to the well-posedness theory, as consequence of the notion of renormalized solutions, in two important situations, namely the Sobolev case (DiPerna and Lions, [56]) and the $BV$ case (Ambrosio, [6]). We present an outline of the theory of renormalized solution and the main points of the DiPerna–Lions regularization strategy in the proof of the well-posedness theory for the continuity equation in the Sobolev case.

1.1 Lagrangian vs Eulerian points of view

In Fluid Dynamics [25], [39], [105], [83], it is customary to consider two complementary ways of describing fluid motion, which are called Lagrangian and Eulerian. Mainly, Eulerian and Lagrangian viewpoints are used as mathematical tools that describe dynamical transport phenomena [75]. Let us profit by this occasion by recall that the most basic facts concerning these two approaches, which will later turn out in Chapter 3 to be very important in the study of the Monge–Kantorovich problem [27], [115]; more can be found in Chapter II of [49]. By the way, we note that the “Lagrangian” point of view was apparently introduced by Euler, while the “Eulerian” point of view should be attributed to Bernoulli and D’Alembert [117].

Let’s start with the Lagrangian point of view, which focuses on the trajectories of particles. In the Lagrangian description, one puts a label on each particle, and then studies the trajectory of each labelled particle. Indeed, when we describe a motion via Lagrangian formulation we give “names” to particles (using either a specific label, or the initial position they had, for instance) and then describe, for every time $t$ and every label, what happens to that particle. “What happens” means providing its position and/or its speed. Hence we could for instance give some functions $X_t(x)$ standing for the position at time $t$ of particle originally located at $x$. More precisely, assuming that we label particles according to their initial position $x_0$, we denote by $x = X_t(x_0)$, the position at time $t$ of a particle that was located at position $x_0$ at time 0. It is usually assumed that for each time $t$, the map $x_0 \mapsto X_t(x_0)$, defined on $\mathbb{R}^d$, is one-to-one. Other possibility, instead of giving names we could consider bundles of particles with the same behavior and indicate how many are they. This amounts to giving a measure on possible behaviors. The description may be more or
less refined. For instance if one only considers two different times \( t = 0 \) and \( t = 1 \), the behavior of a particle is only given by its initial and final positions.

The description may be more or less refined. For instance, if one only considers two different times \( t = 0 \) and \( t = 1 \), the behavior of a particle is only given by its initial and final positions. A measure on those pairs \((x, X_t)\) is exactly a transport plan. This explains why we can consider that the Kantorovitch problem given in Section 3.2 is expressed in Lagrangian coordinates. The Monge problem given in Section 3.1 is also Lagrangian, where particles are labelled by their initial position (see Chapter 3 for more details).

The Lagrangian description may be more or less refined. For instance, if one only considers two different times \( t = 0 \) and \( t = 1 \), the behavior of a particle is only given by its initial and final positions. A measure on those pairs \((x, X_t)\) is exactly a transport plan. This explains how the Kantorovitch’s problem will be given in Section 3.2 is expressed in Lagrangian coordinates. The Monge’s problem will be given in Section 3.1 is also Lagrangian, where particles are labelled by their initial position (see Chapter 3 for more details on such transport problems).

On the other hand, there is an alternative, formally equivalent, way of looking at Fluid Mechanics [57], [62], [66], [82]: the Eulerian point of view, in which the “unknown” mass density of a fluid flow is related to the time–dependent velocity field. In the Eulerian formulation, we describe, for every time \( t \) and every point \( x \), what happens at such a point at such a time. “What happens” usually means what the velocity, the density and/or the flow rate (both in intensity and in direction) of particles located at time \( t \) at point \( x \) are. In the Eulerian formulation, we use two variables, i.e. the density \( \rho(t, x) \) of the fluid and the flow velocity \( v(t, x) \) of the mass elements \( \mu(t, x) \). What is the natural evaluation equation for \( \rho_t \)?

To switch between these two descriptions, it is possible to write the equation satisfied by the mass density of a family of particles moving according to the velocity field \( v \). This means that we prescribe the initial mass density \( \rho(0, \cdot) := \rho_0 \), and that the position of the particle originally located at \( x \) will be given by the solution of the identities

\[
\dot{X}(t, \cdot) = v(t, X(t, \cdot)), \quad \text{under} \quad X(0, \cdot) = \text{Id}.
\]
We define the “time–dependent transport” map $X(t, \cdot) := X_t$ (defined in Chapter 3) through the map $X_t : \mathbb{R}^d \to \mathbb{R}^d$, and we look for the measure $\rho_t := X_t\# \rho_0$ (where $X_t\# \rho_0$ denotes the pushforward of the measure $\rho_0$ via the map $[0, T] \ni t \mapsto X_t$; that is, $\rho_t(B) = \rho_0(X_t^{-1}(B))$ for all Borel sets $B \subset \mathbb{R}^d$). Then we get the Eulerian description of the distribution $X_0\#\mu_0$ as it is transported by the velocity $v$: this means $(\rho, v)$ solves, in the sense of distributions, the so–called Continuity Equation

$$\partial_t \rho + \nabla \cdot (\rho v) = 0, \quad \text{in } [0, T] \times \mathbb{R}^d$$

that is briefly addressed in Sections 1.3 and 1.4. This linear PDE is known in physics as the identity of conservation of mass in the Eulerian formulation, and motivates our notation $\rho$ (reminiscent of the standard notation for the density of a fluid). Thus, in the Eulerian description of Fluid Dynamics, one stares at a given, fixed point of space $x$, and measures the density $\rho(t, x)$ of fluid particles going through this point at time $t$.

The Euler equation is one of the most basic and oldest equations in Fluid Mechanics going back to the eighteenth century, and yet still rich in mysteries [84], [90], [91], [93], [108]. Indeed, it is a basic continuum model for the “fluid” flow. In its simplest version it models an incompressible flow; that is, the divergence of the flow velocity is zero. A key parameter in the qualitative behavior of the flow is the divergence of the vector field $v$. If it is zero, the flow is incompressible; if it is negative, then the flow is contracting and tends to create higher densities as time goes by; if it is positive, then the flow is expanding and tends to create lower densities.

Notice that, if two particles coming from different initial locations occupy the same position at some time $t$ (i.e. $X_t(x_1) = X_t(x_2)$), then there is an ambiguity to remove in the definition above of the velocity field. This is one possible reason for imposing injectivity of the map $X$ (other reasons would rest on physical interpretation). Moreover, in the context of the incompressible continuity equation, it is natural to require that the map $X$ also be surjective: if not, there would be some vacuum created inside the domain, contradicting the fact that the fluid has constant density (incompressibility). This is why it seems natural to search for $\{X(t, \cdot)\}_{t \geq 0}$ as a family of diffeomorphisms from $\mathbb{R}^d$ to $\mathbb{R}^d$. 
Finally, to caricature things a little bit, the Lagrangian formulation is “stronger” than the Eulerian one, but usually breaks down because of the problem of shocks, i.e. the meeting of different trajectories. For most equations of interest in compressible Fluid Mechanics [117], shocks appear in finite time, even for very smooth initial data. However, this is not the case in the problem optimal transportation, as we shall discuss later on. This is why, unlike many other related problems, the Lagrangian problem is best solved with a mass transportation theory. And in fact, our rigorous justification of the Eulerian point of view will be entirely based on the Lagrangian one. Why bother with an Eulerian representation, if it performs less well and cannot be justified rigorously without the help of the Lagrangian machinery? It will turn out that the discussion of the Eulerian point of view is incredibly other problems involving partial differential equations. Therefore, it will be a rewarding investment to investigate it.

1.2 The classical theory of Lagrangian flows

The classical theory of existence and uniqueness of solutions to first–order, scalar differential equations of the general form, as depending on the parameter given by the initial data on the multidimensional Euclidean space, refers to the study of the Cauchy problem for the ordinary differential equation

\[
\begin{align*}
\dot{y}(t) &= v(t, y(t)) \\
y(t_0) &= x_0,
\end{align*}
\] (1.1)

under various regularity assumptions on the (nonautonomous) vector field \( v : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d \) for \( d \geq 1 \); that is, it gives sufficient conditions on \( v \) that guarantee existence (and, maybe, uniqueness) of solutions for some initial data \( y_0 \) and some (small) time. Here we denote with “dot” the differentiation with respect to the time variable \( t \).

If the velocity field \( v : (t, x) \mapsto v_t(x) \in \mathbb{R}^d \) (throughout this work we often use the notation \( v := v_t \)) is a nicely behaved function, the whole theory is simple, classical and well–understood. We want to give characterizations for the well–posedness of (1.1) under the assumption that the vector field \( v \) is sufficiently smooth; that is, Lipschitz continuous
with respect to the spatial variable, uniformly with respect to time. This case has become known as the Cauchy–Lipschitz theory. It basically shows (Picard–Lindelöf Theorem) that, if \( \mathbf{v} \) is a continuous time–dependent vector field defined on an open set \( \mathcal{O} \subset \mathbb{R} \times \mathbb{R}^d \) containing a rectangle

\[
\mathcal{R} = \{ (t, x) \in \mathbb{R} \times \mathbb{R}^d : |t - t_0| \leq a, |x - x_0| \leq b \},
\]

and satisfies the condition

\[
|\mathbf{v}_t(x)| \leq L(1 + |x|), \quad |\mathbf{v}_t(x) - \mathbf{v}_t(\bar{x})| \leq L|x - \bar{x}| \quad \text{for every } (t, x), (t, \bar{x}) \in \mathcal{R} \quad (1.2)
\]

for some constant \( L > 0 \), then a classical solution of the initial value problem (1.1) exists and it is unique for all times \( t \in [t_0 - \tau, t_0 + \tau] \), where \( \tau < \min\{a, (b/L)\} \). That is a function \( \mathbf{y} \in C^1([t_0 - \tau, t_0 + \tau]; \mathbb{R}^d) \) solves the initial value problem (1.1) whenever \( \mathbf{y} \in C([t_0 - \tau, t_0 + \tau]; \mathbb{R}^d) \) satisfies the integral equation

\[
\mathbf{y}(t) = x_0 + \int_{t_0}^{t} \mathbf{v}(s, \mathbf{y}(s)) \, ds \quad \text{for every } t \in [t_0 - \tau, t_0 + \tau].
\]

Notice that, the solution for the Cauchy problem (1.1) requires to fix a point \((t_0, x_0) \in \mathcal{R}\) contained in the open set \( \mathcal{O} \). In general, the (classical) solution of the ODE \( \dot{\mathbf{y}}(t) = \mathbf{v}(t, \mathbf{y}(t)) \) consists of an interval \([t_1, t_2] \subset \mathbb{R}\) and a function \( \mathbf{y} \in C^1([t_1, t_2]; \mathbb{R}^d) \) which satisfies this equation for every \( t \in [t_1, t_2] \). In particular, this implies that \((t, \mathbf{y}(t)) \in \mathcal{O}\) for every \( t \in [t_1, t_2] \). This solution is called an integral curve or a characteristic curve of the vector field \( \mathbf{v} \). Moreover, if \( \mathbf{v} \in C^\infty(\mathcal{O}; \mathbb{R}^d) \), then every characteristic curve of \( \mathbf{v} \) is in \( C^\infty([t_1, t_2]; \mathbb{R}^d) \).

As a consequence, the integral curve \( \mathbf{y}(t) \) is absolutely continuous in \([t_0 - a, t_0 + a]\) and this solution of (1.1) can be uniquely extended until its graph touches the boundary of \( \mathcal{R} \). This identifies a maximal interval of existence for the solution to (1.1). If we consider the solution in the interval \([t_1, t_2]\) with \( t_2 > a \), we can restart \( \mathbf{y} \) at \((a, \mathbf{y}(a))\) until the unique continuous solution arrives at time \( t_2 \). The backward flow is the special case when the time \( t \) is smaller than the initial time. More generally, if a continuous vector field \( \mathbf{v} \) is bounded and globally defined on \( \mathbb{I} \times \mathbb{R}^d \), where \( \mathbb{I} \subset \mathbb{R} \) is an interval, then for every \((t_0, x_0) \in \mathbb{I} \times \mathbb{R}^d\) there exists a unique solution \( \mathbf{y} \in C^1(\mathbb{I}; \mathbb{R}^d) \) to (1.1) is defined for every \( t \in \mathbb{I} \). Furthermore,
looking at the solution of \((1.1)\) as a function not only of the time but also of the initial point \(x_0\) leads to the following definition:

**Definition 1.1 (Flow of a time–dependent vector field)** Let \(v : I \times \mathbb{R}^d \to \mathbb{R}^d\) be a continuous and bounded vector field, where \(I \subset \mathbb{R}\) is an interval. The (classical) flow corresponding to the velocity \(v\) starting at time \(t_0 \in I\) is a map \(X : I \times \mathbb{R}^d \to \mathbb{R}^d\), which gathers together all trajectories, in the sense that it solves the DE

\[
\begin{align*}
\partial_t X(t, x) &= v(t, X(t, x)), \\
X(t_0, x) &= x, \quad x \in \mathbb{R}^d.
\end{align*}
\]

(1.3)

Note first that, the solution formula \(X(t, x)\) is written as a function of two variables in order to account not only for the time variable, but also for the initial value prescribed for \(X\), which is the reason why they are instrumental to the Lagrangian description of particle motion. In the case where the solution exists and is unique for all \(x \in \mathbb{R}^d\), loosely speaking, the function \(X\) is called the *classical flow* of \(v\). The terminology comes primarily from Fluid Dynamics: if a given function \(v\) stands for the velocity of fluid flow, then \(X(t, x)\) accounts for the position at time \(t\) of the fluid particle that was initially \((t = t_0)\) at position \(x\). Then, given \(x \in \mathbb{R}^d\), we define the flow map \(X\) through \(I \ni t \mapsto X(t, \cdot) \in \mathbb{R}^d\) as the value at time \(t\) of the solution of the nonautonomous ODE \((1.1)\); that is, \(X(t, \cdot)\) is characterized by the Cauchy problem \((1.3)\). We are thus looking for characteristic (or integral) curves of the given velocity field \(v\); that is, curves with the property that at each point the tangent vector coincides with the value of the given vector field at such point.

The following proposition immediately deduces a global result in time of the existence and uniqueness of the flow, due to the theorems of Peano and Picard–Lindelöf (who constructed only local solutions in time). This (classical) result requires a (globally defined) continuous and bounded vector field \(v\) is locally Lipschitz continuous with respect to the spatial variable, with some uniformity in the time. Moreover, the Lipschitz regularity of \(X\) can be easily shown by the Gronwall’s lemma. For details we refer to [12].

**Proposition 1.2** Let \(I \subset \mathbb{R}\) be an interval. If \(v_t\) is continuous in \(x\) for every \(t \in I\), then for every initial datum in \((1.3)\), there is local existence (there exists at least a solution, defined
on a neighborhood of \( t = t_0 \). If, moreover, \( \mathbf{v} \) satisfies \( |\mathbf{v}_t(x)| \leq C_0 + C_1|x| \), then the solution is global in time. Furthermore, if \( \mathbf{v}_t \) is Lipschitz in \( x \) and uniformly in \( t \), then the solution is unique and defines a flow \( \mathbf{X}(t, \cdot) : \mathbb{R}^d \to \mathbb{R}^d \). In this case, if we set \( L := \sup_t \text{Lip}(\mathbf{v}_t) \), then the map \( \mathbf{X}(t, \cdot) \) is Lipschitz in \( x \), with constant \( e^{L|t-t_0|} \); finally, \( \mathbf{X}(t, \cdot) \) is invertible and its inverse is also Lipschitz with the same constant.

**Proof.** We do not prove existence, which can be classically obtained by fixed point arguments in the space of curves (see Theorems 1.1.2 and 1.2.4 of [41]). As for uniqueness and for the dependence on the initial datum \( x \in \mathbb{R}^d \), we consider two distinct solutions \( \mathbf{X}_1 \) and \( \mathbf{X}_2 \), to the initial value problem (1.3), and we define \( E(t) := |\mathbf{X}_1(t, x) - \mathbf{X}_2(t, x)|^2 \) and notice that \( E(t_0) = 0 \). We have

\[
\dot{E}(t) = 2 (\mathbf{X}_1(t, x) - \mathbf{X}_2(t, x)) \cdot (\mathbf{v}_t(\mathbf{X}_1(t, x)) - \mathbf{v}_t(\mathbf{X}_2(t, x)))
\leq 2 \text{Lip}(\mathbf{v}_t) (|\mathbf{X}_1(t, x) - \mathbf{X}_2(t, x)|^2 + |\mathbf{X}_1(t, x) - \mathbf{X}_2(t, x)|),
\]

which proves \( |\dot{E}(t)| \leq 2 LE(t) \), where \( \text{Lip}(\mathbf{v}_t) \) denotes, as usual, the Lipschitz constant of \( \mathbf{v}_t \). By Gronwall’s lemma, this gives \( E(t_0)e^{-2L|t-t_0|} \leq E(t) \leq E(t_0)e^{2L|t-t_0|} \) for every \( t \geq t_0 \), and provides at the same time uniqueness, injectivity, and the bi–Lipschitz behavior of \( \mathbf{X}(t, \cdot) \) (which is also a homeomorphism from \( \mathbb{R}^d \) onto \( \mathbb{R}^d \) since for every \( x \in \mathbb{R}^d \) we can solve (1.1) imposing the Cauchy datum \( y(t_0) = x_0 \)).

The proof of the Lipschitz regularity of the flow \( \mathbf{X}(t, x) \) with respect to \( x \) goes along the same lines. Fixed two points \( x, \bar{x} \in \mathbb{R}^d \) and compare the two different solutions to the DE (1.3) with initial data at time \( t = t_0 \); respectively, equal to \( x \) and \( \bar{x} \). We can compute \( \mathbf{X}(t, x) - \mathbf{X}(t, \bar{x}) = x - \bar{x} + \int_{t_0}^t (\mathbf{v}_s(\mathbf{X}(s, x)) - \mathbf{v}_s(\mathbf{X}(s, \bar{x}))) \, ds \). Hence we deduce

\[
|\mathbf{X}(t, x) - \mathbf{X}(t, \bar{x})| \leq |x - \bar{x}| + \text{Lip}(\mathbf{v}_t) \int_{t_0}^t |\mathbf{X}(s, x) - \mathbf{X}(s, \bar{x})| \, ds.
\]

A simple Gronwall argument then gives

\[
|\mathbf{X}(t, x) - \mathbf{X}(t, \bar{x})| \leq |x - \bar{x}| e^{L|t-t_0|}.
\]

This means that the solution depends Lipschitz continuously on the initial data; that is, \( \mathbf{X}(t, x) \) is Lipschitz with respect to \( x \), and the Lipschitz constant depends exponentially on the Lipschitz constant of the given velocity field \( \mathbf{v} \). A very similar argument shows that, if we consider the two different solutions to (1.3) with the same initial data but relative to different vector fields, we have the continuity estimate

\[
|\mathbf{X}(t, x) - \mathbf{X}(t, \bar{x})| \leq |t - t_0| \|v - \bar{v}\|_\infty e^{L|t-t_0|}.
\]

The injectivity can be derived from the uniqueness of the backward flow: If the flow \( X \) starts from two points \( x \neq \bar{x} \) and arrives at some \( t \) at the same point \( \mathbf{X}(t, x) = \mathbf{X}(t, \bar{x}) = \hat{x} \),
the backward flow starting from \((t, \hat{x})\) will be not unique. Regarding the surjectivity: for every point \(x' \in \mathbb{R}^d\) one can find a backward flow \(X(t', x) = x' + \int_{t'}^t v(s(X(s, x))) \, ds = x \in \mathbb{R}^d\) starting from \((t, x')\) according to the classical Cauchy–Lipschitz theory. In case \(t' = t_0\) yields \(X(t, x) = x'\). The Lipschitz regularity of \(X\) is already shown above. Since the Lipschitz continuity gives only the local \(C^1\)–regularity, the \(C^1\)–regularity of \(X(t, \cdot)\) in \(\mathbb{R}^d\) one can follow the results in [12], which states that if \(v\) has \(C^1\)–regularity in space, then the flow \(X(t, \cdot)\) is also \(C^1\) in space, and hence we derive the desired statement. In other words, the flow \(X(t, x)\) inherits the Lipschitz regularity with respect to \(x\).

As mentioned above, if the vector field \(v\) is bounded and Lipschitz with respect to the space variable we immediately deduce existence and uniqueness of the flow from Proposition 1.2. Moreover, Proposition 1.2 also gives Lipschitz regularity of the flow. If we drop the Lipschitz continuity assumption, then we may lose the uniqueness of the solution; however, we still have existence, thanks to Peano’s theorem, where we do not need continuity in both variables for that. A possible example is very well known as the Peano phenomenon for the square root, which stresses that some smoothness assumptions are necessary to obtain uniqueness.

**Example 1.3 (The Peano phenomenon for the square root)** Let us consider in the real line the autonomous vector field \(v_t : \mathbb{R} \to \mathbb{R}\) given by the square root function: \(v_t(y) = 2\sqrt{|y|}\). One can readily check that the one–dimensional differential equation

\[
\partial_t X(t, x) = v_t(X(t, x)) = 2\sqrt{|X(t, x)|} \quad \text{for} \quad t \in [0, T]
\]

with some prescribed initial datum \(X(0, x) = x \in \mathbb{R}\) has two different solutions \(X_1(t, x) = 0\) and \(X_2(t, x) = t^2\) with the same value \((x = 0)\) at the initial time \(t_0 = 0\). Since \(v\) is Hölder continuous but not Lipschitz (due to the singularity at \(y = 0\)), the Cauchy–Lipschitz theory (regarding existence and uniqueness of classical solutions to ODE’s) does not apply, and in fact uniqueness of the trajectories \(X\) fails. Indeed, when \(x = 0\), it is easy to construct an
infinite family of distinct solutions to (1.4), each of the form

\[
X^\varepsilon(t, x) = \begin{cases} 
0 & \text{if } t \in [0, t_\varepsilon], \\
(t - t_\varepsilon) |t - t_\varepsilon| & \text{if } t \in [t_\varepsilon, T]
\end{cases}
\]

for any value of the time parameter \(0 \leq \varepsilon < T\). The solution can “stay at rest” in the origin for an arbitrary time. From an heuristic point of view, what goes wrong in this example is precisely the stopping of the trajectories at the point \(x = 0\). If we consider initial data in the interval \([-1, 0]\), and for any such initial data we pick exactly the trajectory that stays indefinitely at \(x = 0\) once it reaches this point, then we see that the whole interval \([-1, 0]\) collapses to one point in finite time, under the action of this flow. This example will be considered again later: see Example 5.14.

Notice that, the lack of uniqueness in the preceding example is due to the simultaneous occurrence, at the origin, of both the lack of regularity and the vanishing of \(v\). Indeed, for general continuous vector fields \(v\), under the assumption that \(v_t(x) \neq 0\), we can divide \(\partial_t X(t, x) = v_t(X(t, x))\) by \(v_t(X(t, x))\) and integrate in a neighbourhood of \(x\). Thus, \(X(t, x)\) must be locally given by an explicit formula involving the inverse of the primitive of \(1/v_t\), and uniqueness follows. This is only true in a neighbourhood of \(t = t_0\): similarly to the case described in Example 2.9, the solution of the problem (1.4) with any initial data \(x_0 < 0\) reaches the origin (and up to that time it is unique), but it can stay at rest there for an arbitrary amount of time, before leaving it. This means that (while uniqueness does not hold in general) the continuity of \(v\) is enough to guarantee (local) existence of a solution, due to the so–called Peano theorem. We refer for instance to [12], [43], [74] for an introduction to the classical theory of ODE’s and for further remarks on this example.

**Remark 1.4** Uniqueness result for (1.1) also holds under the following Osgood condition [41] (see also [32], [33] for recent results on the transport equation when \(b\) satisfies a one-sided Lipschitz condition): \(|v_t(x) - v_t(\bar{x})| \leq \omega(|x - \bar{x}|)\) for every \((t, x), (t, \bar{x}) \in \mathbb{R} \times \mathbb{R}^d\), where \(\omega: \mathbb{R}_+ \to \mathbb{R}_+\) is an increasing function satisfying \(\omega(0) = 0\), \(\omega(z) > 0\) for every \(z > 0\) and \(\int_0^1 1/\omega(z) \, dz = +\infty\). Moreover, uniqueness forward in time can be given under the following one–sided Lipschitz condition: \((v_t(x) - v_t(\bar{x}))(x - \bar{x}) \leq L|x - \bar{x}|^2\) for every \((t, x, \bar{x}) \in \mathbb{R}^{1+2d}\).
By “forward in time” we mean the following: two solutions \( X_1 \) and \( X_2 \) to (1.3), which by definition satisfy \( X_1(0, x) = X_2(0, x) = x \), coincide for \( t > t_0 \). Observe that the one-sided Lipschitz condition cannot guarantees uniqueness backward in time for the flow \( X \) of the vector field \( v \), since it is not invariant under an inversion of the sign of \( t \) in the equation, which amounts to a change of sign of \( v \).

We now present other conditions which are a bit more general than the Lipschitz condition required in Proposition 1.2, but nevertheless they are sufficient to get uniqueness. In the smooth setting, one can easily identify a unique flow map \( X \) of the vector field \( v \). If \( v \) is not smooth enough to justify Proposition 1.2 above (since it may be discontinuous in time), then the proposition above does not apply and the Cauchy problem (1.3) may not admit any solution. However, the result in Proposition 1.2 can be still valid even if \( v \) is not continuous in time provided we assume that \( v \) is only summable with respect to the time variable and Lipschitz continuous in \( x \) uniformly in \( t \). Relaxing the assumptions on the time variable \( t \) of a vector field \( v \) to be integrable in the interval \( I \subset \mathbb{R} \) refer to the following Carathéodory theorem [74]:

**Theorem 1.5** Let \( v_t \) be locally bounded by a nonnegative integrable function defined on \( I \), where \( I \subset \mathbb{R} \) is an interval. If the map \( t \mapsto v_t(x) \) is measurable in \( I \) for every \( x \in \mathbb{R}^d \) and the map \( x \mapsto v_t(x) \) is continuous for each \( t \in I \), then the Cauchy problem (1.1) has at least one solution. Moreover, if the map \( x \mapsto v_t(x) \) is Lipschitz continuous for each \( t \in I \), with a uniform Lipschitz constant, then the Cauchy problem (1.3) has a unique solution, depending Lipschitz continuously on the initial data \( x \).

For a direct proof of this theorem see for instance Theorem 1 of [65]. As a consequence of the proof, the flow \( X(t, x) \) is absolutely continuous in an interval \( I \subset \mathbb{R} \). By a solution of (1.3) we mean an absolutely continuous function \( X : I \times \mathbb{R}^d \rightarrow \mathbb{R}^d \) satisfying the integral solution

\[
X(t, x) = x + \int_{t_0}^{t} v_s(X(s, x)) \, ds \quad \text{for} \quad \mathcal{L}^d\text{-a.e.} \ x \in \mathbb{R}^d \text{ and every } t \in I. \tag{1.5}
\]

A helpful property of the flow \( X \) will be given in the following corollary.
Corollary 1.6 For every \( t \in \mathbb{I} \subset \mathbb{R} \), the mapping \( X(t, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d \) is Lipschitz continuous and a diffeomorphism.

As we already observed, if \( \mathbf{v} \) is sufficiently regular, then the flow map \( X_t \) is a diffeomorphism of \( \mathbb{R}^d \). More generally, if \( \mathbf{v} \in C^1_c(\mathbb{I} \times \Omega) \), where \( \Omega \subset \mathbb{R}^d \) is open, then the unique flow map \( X_t \) is a diffeomorphism of \( \Omega \) \([95], [113]\). This means that, thanks to Proposition 1.2: for instance, if \( \mathbb{I} = [0, T] \) and a vector field \( \mathbf{v} = \mathbf{v}(t, x) \) is smooth enough, then \([0, T_0) \ni t \mapsto X_t \) is a locally Lipschitz in \( \mathbb{R}^d \), with \( X_0 = \text{Id} \). Moreover, the flow map \( X_t : \mathbb{R}^d \rightarrow \mathbb{R}^d \) is a bijection for every \( t \in [0, T_0) \), and for every \( T < T_0 \) and for every compactly supported \( \Omega \subset \mathbb{R}^d \), the maps \((t, x) \mapsto X_t(x)\) and \( X_t^{-1}(x) \) are Lipschitz on \( \mathbb{I} \times \Omega \). Consequently, if \( \mathbf{v} \) is given and uniformly Lipschitz on \( \mathbb{R}_+ \times \mathbb{R}^d \), then, thanks to Corollary 1.6, it generates a family of (globally) Lipschitz diffeomorphisms for all times. In this case, the vector field \( \mathbf{v} \) is called a velocity associated to the flow \( X \).

1.3 The Continuity Equation in the smooth setting

After recalling some basic properties of the flow, we try now, in this section, to identify the solutions of the continuity equation in the higher dimensional setting, and we want to connect them to the flow of the vector field \( \mathbf{v}_t \). Assume now \( T \in (0, +\infty) \) and we are given a probability measure \( \mu_0 \), that captures a certain initial spatial distribution of particles that follow the flow of \( \mathbf{v} \), a natural question is: how does this spatial distribution of particles evolve with time? In other words, how is the initial distribution \( \mu_0 \) transported by the flow of \( \mathbf{v} \)? In transport terms, this amounts to characterize the curve of probability measures \((0, T) \ni t \mapsto X_{t\#}\mu_0 \). The expression \( X_{t\#}\mu_0 \) denotes the pushforward of the initial measure \( \mu_0 \) through the map \( X_t : \mathbb{R}^d \rightarrow \mathbb{R}^d \), defined according to Definition 3.2. We shall see that \( X_{t\#}\mu_0 \) is characterized by the following Cauchy problem on \((0, T) \times \mathbb{R}^d \):

\[
\begin{cases}
\partial_t \mu(t, \cdot) + \nabla \cdot (\mathbf{v}(t, \cdot) \mu(t, \cdot)) = 0, \\
\mu(0, \cdot) = \mu_0,
\end{cases}
\] (1.6)
where $\mu(t, \cdot) := \{\mu_t\}_{t \in (0, T)}$ is a family of Borel probability measures on $\mathbb{R}^d$, which depends on the time parameter $t \in (0, T)$. Thus, one can see obviously that, in the classical situation, there is a strong connection between the ODE (1.3) and the continuity equation, an evolutionary partial differential equation which has the form (1.6) above, from Fluid Dynamics.

Note that since we did not assume any regularity on $\mu_0$ yet ($\mu_0$ could be a Dirac mass and then $X_t \# \mu_0$ would remain a Dirac mass for every $t \in (0, T)$), we have to understand the continuity equation in some appropriate weak sense, i.e. in the sense of distributions.

### 1.3.1 Weak and Measure–Valued Solutions

We first introduce the weak formulation for the continuity equation (1.6). Let $v : (t, x) \mapsto v_t(x) \in \mathbb{R}^d$ be a Borel vector field satisfying the following condition:

\[
v_t \in L^1(\mu_t; \mathbb{R}^d) \text{ for } L^1\text{-a.e. } t \in (0, T) \quad \text{and} \quad (0, T) \ni t \mapsto \|v_t\|_{L^1(\mu_t; \mathbb{R}^d)} \in L^1(0, T),
\]

and denote by $\nabla \cdot v = \sum_{j=1}^d \partial_x v_j$ the divergence of $v_t$ with respect to the spatial variable in the sense of distributions. The continuity equation is intended in distributional sense, according to the following definition.

**Definition 1.7** If a vector field $v : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying the condition (1.7), then a family $(0, T) \ni t \mapsto \mu_t$ of probability measures on $\mathbb{R}^d$ is said to be a measure–valued (or, a weak) solution of (1.6) if the following hold:

(i) it is continuous in the sense that for every $\xi \in C^1_c(\mathbb{R}^d)$, the map

\[
t \mapsto \langle \xi, \mu_t \rangle := \int_{\mathbb{R}^d} \xi \, d\mu_t \text{ is continuous on } (0, T) \quad \text{and} \quad \langle \xi, \mu_0 \rangle = \int_{\mathbb{R}^d} \xi \, d\mu_0;
\]

(ii) for every bounded and Lipschitz test function $\xi \in C^1_c([0, T] \times \mathbb{R}^d)$, one has

\[
\int_0^T \int_{\mathbb{R}^d} \left( \partial_t \xi(t, x) + v_t(x) \cdot \nabla \xi(t, x) \right) \, d\mu_t(x) \, dt = -\int_{\mathbb{R}^d} \xi(0, x) \, d\mu_0(x).
\]

Notice that this definition is the standard notion of weak solution of a PDE. We can equivalently define (i) above noticing that, the mapping $t \mapsto \langle \xi, \mu_t \rangle$ is Lipschitz on $(0, T)$. 
For smooth solutions, equation (1.9) can be immediately deduced from (1.6) multiplying it by \( \xi \) and integrating by parts, where the divergence operator \( \nabla \cdot (v_t \mu_t) \) is defined by duality, through the formula
\[
\int \xi \, d((\nabla \cdot (v_t \mu_t))) = - \int \nabla \xi \cdot v_t \, d\mu_t,
\]
where \( \xi \) is a smooth test function with compact support.

We can redefine the formula (1.9) with a vector field \( v : (0, T) \times \Omega \to \Omega \) satisfying
\[
v_t \in L^1(\mu_t; \Omega) \quad \text{and} \quad \int_0^T \|v_t\|_{L^1(\mu_t; \Omega)} \, dt = \int_0^T \int_\Omega |v_t(x)| \, dx \, dt < +\infty.
\]
Here \( \Omega \subset \mathbb{R}^d \) is a “smooth” bounded domain in \( \mathbb{R}^d \). In this case, we take \( \xi \in C^1_c([0, T] \times \Omega) \) and require the support to be far from \( t = T \) but not from \( \partial \Omega \), when \( \Omega \) is bounded; also \( \Omega \) is usually supposed to be itself closed, but we write \( \overline{\Omega} \) to stress the fact that we do include its boundary. Obviously the new formulation will include Neumann “no-flux” boundary conditions on \( \partial \Omega \) for \( v_t \) (if \( \Omega \) is not \( \mathbb{R}^d \) itself), i.e. the tangential condition
\[
v_t(x) \cdot n(x) = 0 \quad \text{for every} \quad (t, x) \in (0, T) \times \partial \Omega
\]
is satisfied, where \( n(x) \) denotes the exterior normal vector to \( \Omega \) at \( x \in \partial \Omega \), which implies \( \Omega \) is invariant by the flow (trajectories starting in \( \overline{\Omega} \) stay in \( \overline{\Omega} \) forever). If we want to impose the initial and final measures, we can say that a family of pairs \( (\mu, v) \) of measures \( \mu_t \) and vector fields \( v_t \) solves together the same continuity equation, in the sense of distribution, with initial and final data \( \mu_0 \) and \( \mu_T \), respectively, if for any test function \( \xi \in C^1_c([0, T] \times \overline{\Omega}) \) (now we do not require the support to be far from \( t = 0, T \)), we have
\[
\int_0^T \int_\Omega \left( \partial_t \xi + v_t \cdot \nabla \xi \right) \, d\mu_t \, dt = \int_\Omega \xi(T, x) \, d\mu_T(x) - \int_\Omega \xi(0, x) \, d\mu_0(x).
\]
(1.10)

Note that, the continuity equation (1.6) is imposed in \( \mathbb{R}^d \) and \( \mu_t(\mathbb{R}^d \setminus \Omega) = 0 \) for every \( t \in (0, T) \) with \( T > 0 \) and every open subset \( \Omega \) of \( \mathbb{R}^d \).

On the other hand, a weak solution of the continuity equation can be defined through the following condition: we say that \( (\mu_t, v_t) \) solves (1.6) in the weak sense if for any test function \( \xi \in C^1_c(\overline{\Omega}) \), the function \( t \mapsto \int_\Omega \xi \, d\mu_t \) is absolutely continuous in \( t \) (In particular, this map belongs to \( W^{1,1}(0, T) \) for every \( \xi \in C^1_c(\overline{\Omega}) \)), and that its distributional derivative is
\[
\frac{d}{dt} \int_\Omega \xi \, d\mu_t = - \int_\Omega \xi \, d[\nabla \cdot (v_t \mu_t)] = \int_\Omega \nabla \xi(x) \cdot v_t(x) \, d\mu_t(x) \quad \text{for a.e.} \quad t \in (0, T).
\]
(1.11)

Note that, in this case \( t \mapsto \mu_t \) is automatically continuous for the weak convergence, and imposing the values of \( \mu_0 \) and \( \mu_T \) may be done pointwisely in time. The following standard
result in the theory of evolutionary PDE’s shows the regularity in time of the measure-valued solutions \( t \mapsto \mu_t \) of the problem (1.6) in \( \mathcal{P}(\mathbb{R}^d) \) (the space of Borel probability measures on \( \mathbb{R}^d \)).

**Proposition 1.8 (Weak continuity in time)** Let \( \mu_t \) be a Borel family of probability measures satisfying (1.9) for a Borel vector field \( \mathbf{v}_t \) satisfying (1.7). Then, every distributional solution admits a representative (another family \( (0, T) \ni t \mapsto \tilde{\mu}_t \in \mathcal{P}(\mathbb{R}^d) \) for \( L^1 \)-a.e. \( t \in (0, T) \)) which is weakly continuous in time.

**Proof.** First, note that by a simple density argument we can find a unique representative \( \tilde{\mu}_t \) independent of \( \xi \), such that \( (0, T) \ni t \mapsto \langle \xi, \tilde{\mu}_t \rangle \) is uniformly continuous in \( (0, T) \) for any \( \xi \in C^1_c(\mathbb{R}^d) \). Indeed, the distributional derivative (1.11) implies in particular, for every \( \xi \) with \( \text{spt}(\xi) \subset B_R(0) \), the estimate

\[
\left| \frac{d}{dt} \langle \xi, \mu_t \rangle \right| = \max_{\mathbb{R}^d} |\nabla \xi| \beta_R(t), \quad \text{where} \quad \beta_R(t) := \int_{B_R(0)} |\mathbf{v}_t(x)| d\mu_t(x) \in L^1(0, T). \tag{1.12}
\]

Let \( L_\xi \subset (0, T) \) be the set of the Lebesgue points of the map \( (0, T) \ni t \mapsto \langle \xi, \mu_t \rangle \). We know that \( L^1((0, T) \setminus L_\xi) = 0 \). We consider for every \( R \in \mathbb{N} \) a countable set \( Z_R \) which is dense in \( C^1_c(B_R(0)) \) with respect to the usual \( C^1 \) norm \( \| \xi \|_{C^1} := \max_{\mathbb{R}} \{ |\xi|, |\nabla \xi| \} \) and set \( L_{Z_R} = \cap_{\xi \in Z_R} L_\xi \). Clearly we have \( L^1((0, T) \setminus L_{Z_R}) = 0 \). The restriction of the curve \( \mu_t \) to \( L_{Z_R} \) provides a uniformly continuous family of bounded functionals on \( C^1_c(B_R(0)) \), since estimate (1.12) implies \( \| \langle \xi, \mu_t \rangle - \langle \xi, \mu_s \rangle \| \leq \| \xi \|_{C^1} \int_s^t \beta_R(\tau) d\tau \) for every \( s, t \in L_{Z_R} \). Therefore \( \mu := \{ \mu_t \}_{t \in L_{Z_R}} \) can be extended in a unique way to a continuous curve \( (0, T) \ni t \mapsto \tilde{\mu}_t^R \in [C^1_c(B_R(0))]' \). Applying iteratively this argument a countable number of times, we can construct in a unique way a continuous curve \( \mu := \{ \tilde{\mu}_t \}_{t \in (0, T)} \in [C^1_c(B_R(0))]' \). Recalling that \( L_{Z_k} \ni t \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d) \) is tight, this also implies that \( \tilde{\mu}_t \) is a continuous curve in \( \mathcal{P}(\mathbb{R}^d) \), for \( t \in (0, T) \).

**Remark 1.9** From the argument of the proof given above, which is indeed classical in the theory of evolutionary PDE’s (see for instance Lemma 8.1.2 of [17] or Theorem 4.1.1 of [49]), we deduce that, up to a modification of \( \mu(t, \cdot) \) in a negligible set of times, we can assume that \( (0, T) \ni t \mapsto \mu(t, \cdot) \) is weakly*-continuous from \( (0, T) \) into \( \mathcal{P}(\mathbb{R}^d) \). It follows in particular
that $\mu(t, \cdot)$ is defined for every $t \in (0, T)$, and in particular at the endpoints; this also gives a sense to the Cauchy data at $t = 0, T$.

**Lemma 1.10** The two notions of weak and distribution solutions to the continuity equation (1.6) are indeed equivalent.

**Proof.** First, let us take the distributional solution (1.10) and test it against test function of the form $\zeta(t)\xi(x)$ with $\zeta \in C^1_{c}([0, T])$ and $\xi \in C^{1}_{c}(\mathbb{R}^{d})$, that is

$$- \int_{0}^{T} \zeta'(t) \langle \xi, \mu_{t} \rangle \, dt = \int_{0}^{T} \zeta(t) \int_{\mathbb{R}^{d}} \nabla \xi(x) \cdot v_{t}(x) \, d\mu_{t}(x) \, dt,$$

so we easily obtain that the map $(0, T) \ni t \mapsto \langle \xi, \mu_{t} \rangle$ belongs to $W^{1,1}(0, T)$ and that (1.11) holds in the sense of distributions over $(0, T)$. Indeed, the arbitrariness of $\zeta$ in (1.13) shows that the distributional derivative (in time) of the map $\langle \xi, \mu_{t} \rangle$ is $\langle \nabla \xi, v_{t} \mu_{t} \rangle := \int_{\mathbb{R}^{d}} \nabla \xi(x) \cdot v_{t}(x) \, d\mu_{t}(x)$. This last function is $L^{1}$ in time since the condition (1.7) gives

$$\int_{0}^{T} |\langle \nabla \xi, v_{t} \mu_{t} \rangle| \, dt \leq \text{Lip}(\xi) \int_{0}^{T} \|v_{t}\|_{L^{1}(\mu_{t}; \mathbb{R}^{d})} \, dt < +\infty.$$

This implies that $\frac{d}{dt} \langle \xi, \mu_{t} \rangle = \langle \nabla \xi, v_{t} \mu_{t} \rangle \in L^{1}(0, T)$ and hence $(\mu, v)$ is a weak solution.

Conversely, the same computations shows that weak solution satisfy (1.9) for any test function of the form $\zeta(t)\xi(x)$. It is then enough to prove that finite linear combination of these functions are dense in $C^{1}([0, T] \times \Omega)$ for every compact set $\Omega \subseteq \mathbb{R}^{d}$ (to do that we only need to show that the set of polynomial functions is dense for the $C^{1}$ norm in $C^{1}([0, 1]^{d})$).

### 1.3.2 Existence and uniqueness of smooth solutions

After checking the relations between the two different notions of solutions, from now on we will often say “solution” to mean, indistinguishably, “weak solution” or “solution in the distributional sense”. We would like now to give at least an existence and uniqueness result on $\mu$ if $v$ is sufficiently smooth. Indeed, it is evident and classical that smooth functions satisfy the continuity equation (1.6) in the classical sense if and only if they are weak (or distributional) solutions. The main way to produce solutions to the continuity equation is
to use the flow of the vector field \( v_t \). First, let us recall in the following a simple change-of-variable computation and image measures.

**Proposition 1.11** Suppose that \( \rho_0 \in L^1(\mathbb{R}^d) \) is a positive density on \( \Omega \subset \mathbb{R}^d \) and \( X_t : \Omega \rightarrow \mathbb{R}^d \) be a Lipschitz injective map, which is thus differentiable a.e. We suppose that \( \det(\nabla X_t) \neq 0 \) a.e. on \( \{ \rho_0 > 0 \} \). Then, the image measure \( X_t#\rho_0 \) is absolutely continuous and its density \( \rho_t \) is given by the following explicit expression

\[
\rho_t(y) = \frac{\rho_0(X_t^{-1}(y))}{\det(\nabla X_t(X_t^{-1}(y)))}.
\]

(1.14)

By “Lipschitz injective map” we mean the following: A map \( X_t : \mathbb{R}^d \rightarrow \mathbb{R}^d \) such that \( \mathcal{L}^d(N) = 0 \) implies \( \mathcal{L}^d(X_t(N)) = 0 \), i.e. \( X_t \) maps negligible sets into negligible sets.

**Proof of Proposition 1.11.** The proof is a direct consequence of (1.6) and of the classical \( d \)-dimensional area formula [12]:

\[
\int_A \xi |\det(\nabla X_t)| \, dx = \int_{\mathbb{R}^d} \sum_{x \in A \cap X_t^{-1}(y)} \xi(x) \, dy \quad \text{for any Borel set } A \subset \mathbb{R}^d,
\]

which is effectively the change of variables between the measure spaces with Lebesgue measure, where \( \nabla X_t \) denotes the Jacobian of \( X \). For a derivation of the equality (1.14) we set \( A = X_t^{-1}(B) \) and \( \xi = \rho_0 / |\det(\nabla X_t)| \) to obtain

\[
\int_{X_t^{-1}(B)} \rho_0(x) \, dx = \int_B \frac{\rho_0}{|\det(\nabla X_t)|} \circ X_t^{-1}(y) \, dy,
\]

for every Borel set \( B \subset \mathbb{R}^d \), which implies the formula (1.14) for the pushforward of \( |\det(\nabla X_t)| \, dx \) under \( X_t \) is \( dy \).

Note that when \( X_t \) is noninjective, the formula becomes \( X_t#\rho_0 = \rho_t \mathcal{L}^d \) with \( \rho_t \) given by

\[
\rho_t(y) = \sum_{x : X_t(x) = y} \frac{\rho_0(x)}{\det(\nabla X_t(x))}.
\]

The same formulas stay true if \( X_t \) is not Lipschitz, but if there are measurable sets \( E_i \) such that \( |\Omega \setminus \bigcup_i E_i| = 0 \) and \( X_t \) is Lipschitz continuous on each set of a countable family \( E_i \) (i.e. \( X_t \) is only countably Lipschitz functions), with the differential \( \nabla X_t \) which is actually the
differential of the restriction of $X_t$ to each set where it is Lipschitz continuous (and coincides thus with the approximate differential of $X_t$).

We can now check that the result in the smooth framework may be applied to the solution produced by the flow, which is actually smooth, thanks to Proposition 1.11 that allowing to reconstruct its density.

**Proposition 1.12** If $\rho_0$ and $v$ are smooth, then $\rho_t$ is smooth in $t$ and $x$. If $\rho_0$, $v$ and $\nabla \cdot v$ are Lipschitz continuous, then $\rho_t$ is Lipschitz continuous in $t$ and $x$.

**Proof.** If $v_t$ is Lipschitz, the flow map is injective (as a well-known consequence of the uniqueness of the solution of the ODE (1.1)). Hence, the density of the image measure is obtained from the initial density through a simple change of variable involving the Jacobian factor. This means that the regularity of $\rho(t, x)$ only depends on the regularity of the Jacobian $J(t, x) := \det(A(t, x))$ and $A(t, x) = D_x X_t(x)$ where $X_t(x) := X(t, x)$ is defined through (1.3). Notice that we have $A(0, x) = \text{Id}, J(0, x) = 1$ and

$$\dot{A}(t, x) = \partial_t(D_x X_t(x)) = D_x(\partial_t X_t(x)) = D_x(v_t(X_t(x))) = Dv_t(X_t(x)) \cdot A(t, x), \quad (1.15)$$

which implies, thanks to usual matrix calculus; that the differential of the determinant application is given by

$$\dot{J}(t, x) = J(t, x) \text{Tr}(A^{-1}(t, x) \dot{A}(t, x)) = J(t, x) \text{Tr}(A^{-1}(t, x) Dv_t(X_t(x)) A(t, x))$$
$$= J(t, x) \text{Tr}(Dv_t(X_t(x))) = J(t, x) \nabla y \cdot v_t(X_t(x)),$$

where Tr denotes the trace operator on matrices. By letting $G(t, x) := \nabla_y \cdot v_t(X_t(x))$ and assuming $X$ is known, one can deal with $\dot{J}(t, x) = G(t, x) J(t, x)$ as if it were a scalar linear equation in $J$, with $x$ being just a parameter. The solution procedure shows that $J(t, x) = \exp\left(\int_0^t G(s, x) \, ds\right)$ which implies that $J(t, x) > 0$ for every $t \in (0, T)$. This also means that if $G$ is bounded from below, then $J(t, x)$ never vanishes, and if $G$ is smooth in $x$, so is $J(t, x)$. The considerations in Proposition 1.11 above allow to deduce the regularity of $\mu(t, x)$.

From the above proof, one immediately sees that the *incompressibility constraint* can be recast in terms of the flow $X$: if $v$ is regular enough (say $C^1$), then $\nabla \cdot v = 0$ if and
only if $J(t, x) \equiv 1$. Indeed, the identity $\det(D_x X_t(x)) \equiv 1$ is obviously satisfied at time 0, since $X_0(x) = x$ is the identity map ($X(0, \cdot) = \text{Id}$); then the equivalence is a consequence of the identity $\partial_t(\ln J(t, x)) = \nabla \cdot v_t(X_t(x))$. Moreover, one can show easily that $\dot{J}(t, x) \leq \|\nabla \cdot v_t\|_\infty J(t, x)$ gives the following estimates on $J$

$$
\exp \left( - \int_0^t \| [G(s, \cdot)]^- \|_\infty ds \right) \leq J(t, x) \leq \exp \left( \int_0^t \| [G(s, \cdot)]^+ \|_\infty ds \right)
$$

for $\mathcal{L}^d$–a.e. $x \in \mathbb{R}^d$, thanks to Gronwall’s lemma, where $[G]^+$, $[G]^-$ denote the positive and negative parts of the divergence $[\nabla \cdot v]^+$, $[\nabla \cdot v]^-$, respectively. In fact, the lower estimate can be obtained by applying the upper one in a time reversed situation (see [12], [13] for more details).

**Lemma 1.13** Suppose that $\rho$ is Lipschitz continuous in $(t, x)$, that $v$ is Lipschitz in $x$, and that the continuity equation (1.6) is satisfied in the weak sense. Then, the equation is also satisfied in the a.e. sense.

**Proof.** First we note that with our assumptions, both $\partial_t \rho$ and $\nabla \cdot (\rho v)$ are well defined a.e. Fix a countable set $D \subset C^1_c(\bar{\Omega})$ which is dense for the uniform convergence in $C^0_c(\bar{\Omega})$ (for instance, use polynomial functions with rational coefficients times suitable cut–off functions). Fix $t_0$ such that $(t, x) \mapsto \rho(t, x)$ is differentiable at $(t_0, x)$ for a.e. $x$, and also such that $t \mapsto \int_\Omega \nabla \xi \cdot v_t \rho_t$ is differentiable at $t = t_0$ with derivative given by $\int_\Omega \nabla \xi \cdot v_t \rho_t$ for all $\xi \in D$. Almost all $t_0$ satisfy these conditions. Then we can also write, by differentiating under the integral sign,

$$
\frac{d}{dt_{t=t_0}} \int_\Omega \xi \, d\rho_t = \int_\Omega \xi (\partial_t \rho)_{t_0} \, dx
$$

which proves $\int_\Omega (\nabla \xi \cdot v_{t_0}) \rho_{t_0} \, dx = \int_\Omega \xi (\partial_t \rho)_{t_0} \, dx$. Yet, we can write $\int_\Omega \nabla \xi \cdot v_{t_0} \rho_{t_0} \, dx$ as $-\int_\Omega \xi \nabla \cdot (\rho_{t_0} v_{t_0}) \, dx$ (by integration by parts of Lipschitz functions, with no boundary term because of the compact support of $\xi$), and finally we get

$$
\int_\Omega \xi (\partial_t \rho)_{t_0} + \nabla \cdot (\rho_{t_0} v_{t_0}) \, dx = 0 \quad \text{for every } \xi \in D
$$

which is enough to prove $\partial_t \rho_{t_0} + \partial_x (\rho_{t_0} v_{t_0}) = 0$ a.e. in $x$. \qed
We would like now to give a uniqueness result on \( \rho \) in the smooth framework. This is also true in a very general framework (see [17], Proposition 8.1.7) for a proof of the fact that the solution in the space of measures is unique for given \( \mathbf{v} \), but we prefer to give an easier proof which requires to consider smooth (Lipschitz in time and space, which is enough to consider the equation in an a.e. pointwise sense) solutions.

**Theorem 1.14** Suppose that \( \mathbf{v} : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d \) is Lipschitz continuous and bounded in \( x \), uniformly in \( t \), and consider two locally Lipschitz solutions \( \rho_1 \) and \( \rho_2 \) of (1.6); with \( \rho_1(0, \cdot) = \rho_2(0, \cdot) \). Then, \( \rho_1 = \rho_2 \).

*Proof.* The equation (1.6) being linear, we only need to consider a solution \( \rho = \rho_1 - \rho_2 \) with \( \rho_0(x) = 0 \) and prove that it vanishes for every time. Let \( R > 0 \) and let \( B_R(0) := \{ x : |x| < R \} \) be the open ball centered at zero and consider the “energy” \( E_R(\rho_t) = \frac{1}{2} \int_{B_R(0)} \rho_t^2(x) \, dx \). We have

\[
E_R(\rho_t) = \int_{B_R(0)} \rho_t(\partial_t \rho_t) = -\int_{B_R(0)} \nabla \rho_t \cdot \mathbf{v}_t \, dx + \int_{\partial B_R(0)} \rho_t^2 \mathbf{v}_t \cdot \mathbf{n} \, dx
\]

\[
= \int_{B_R(0)} \frac{1}{2} \rho_t^2 \nabla \cdot \mathbf{v}_t \, dx + \int_{\partial B_R(0)} \rho_t^2 \mathbf{v}_t \cdot \mathbf{n} \, dx \leq C E_R(\rho_t) + C \partial E_R(\rho_t),
\]

where we used uniform bounds on both \( \nabla \cdot \mathbf{v}_t \) and \( \mathbf{v}_t \). Then, we fix \((t_0, R_0)\) and we apply Gronwall Lemma to \( t \mapsto f(t) := E_{R_t}(\rho_t) \) where \( R_t := R_0 + C t_0 - C t \), then getting

\[
E_{R_{t_0}}(\rho_{t_0}) = f(t_0) \leq f(0) e^{C t_0} = E_{R_0}(\rho_0) e^{C t_0} = 0.
\]

Thus, a simple application for Gronwall’s lemma allows to prove \( E_R(\rho_t) = 0 \) for every \( t \), since \( E_R(\rho_0) = 0 \), and gives the desired thesis. \( \square \)

We can also give a variant of this theorem to adapt to the case of bounded domains. The result is not sharp, as global bounds on \( \mathbf{v} \) and on its Lipschitz constants are not really necessary, and we refer the reader to [17] for a more general proof. Assume \( \Omega \subset \mathbb{R}^d \) is a compact domain and \( \mathbf{v} : (0, T) \times \Omega \to \mathbb{R}^d \) is smooth. If we consider \( E(\rho_t) = \frac{1}{2} \int_{\Omega} \rho_t^2(x) \, dx \), then we have

\[
\dot{E}(\rho_t) = \int_{\Omega} \rho_t(\partial_t \rho_t) = \int_{\Omega} \nabla \rho_t \cdot \mathbf{v}_t \rho_t = \int_{\Omega} \nabla \frac{1}{2} \rho_t^2 \cdot \mathbf{v}_t = -\int_{\Omega} \frac{1}{2} \rho_t^2 \nabla \cdot \mathbf{v}_t \leq C E(\rho_t),
\]

where we used \( -\nabla \cdot \mathbf{v}_t \leq C \) as a consequence of the Lipschitz continuity assumption. This implies that \( \frac{d}{dt} \left[ E(\rho_t) e^{-C t} \right] \leq 0 \), hence \( E(\rho_t) e^{-C t} \leq E(\rho_0) = 0 \) for every \( t \in (0, T) \), which gives the thesis.
The conclusion of previous results on the continuity equation is the following corollary, which can be proven by putting together the results of Lemma (1.13) and Theorem (1.14).

**Corollary 1.15** Suppose that $\rho_0$, $v$ and $\nabla \cdot v$ are Lipschitz continuous functions of $x \in \mathbb{R}^d$ uniformly in $t \in [0, T]$, and that $v$ is uniformly bounded. Then there exists a unique Lipschitz continuous solution to the continuity equation (1.6) (with Neumann condition and initial datum $\rho_0$) and it is obtained through the flow of $v_t$.

### 1.3.3 Connection to the Transport Equation

It is worth mentioning that the flow equation (1.3) and the continuity equation (1.6) are closely connected to the transport equation, a first–order inhomogeneous linear evolutionary PDE with a source term $f \in C^1$ on the right hand side, which has the form

$$
\partial_t F(t, x) + v(t, x) \cdot \nabla F(t, x) = f(t, x), \quad (t, x) \in (0, T) \times \mathbb{R}^d.
$$

(1.16)

Notice that, this equation reduces to the continuity equation when $f_t = -F_t \nabla \cdot v_t$. The two equations are somehow dual: if a vector field $v$ is fixed and we take smooth solutions $\mu_t$ of (1.6) and $F_t$ of (1.16), then one can establish formally a duality between them via the identity

$$
\frac{d}{dt} \int F_t d\mu_t = \int (\partial_t F_t) \mu_t + \int F_t (\partial_t \mu_t) = \int (-v_t \cdot \nabla F_t + f) d\mu_t + \int v_t \cdot \nabla F_t d\mu_t = \int f d\mu_t.
$$

This duality method is a classical tool to prove uniqueness (see the proof of Theorem 1.18 below) in a sufficiently smooth setting (but see also [33], [32]).

We also observe the case when $f \equiv 0$ (this can easily be seen by imposing $\nabla \cdot v = 0$), the both equations coincide. In this case, we have the homogeneous transport equation

$$
\partial_t F_t + v_t \cdot \nabla F_t = 0, \quad F_t : \mathbb{R}^d \to \mathbb{R} \text{ for every } t \in (0, T).
$$

(1.17)

It is important to keep in mind that this equation represents the *incompressible* Eulerian expression of the Lagrangian flow. The condition $\nabla \cdot v = 0$ means that the fluid is incompressible. A vector field $v$ satisfying this incompressibility condition is said to be *divergence-free.*
Proposition 1.16 Suppose that, $v_t$ is a smooth vector field (say, $C^1$, we will see later the general case of the vector field), with $X_t$ being its flow map (with, as usual, characterized by (1.3)). Also suppose that, the equation (1.17) coupled with an initial data $F(0, \cdot) = F_0$ in $C^1(\mathbb{R}^d)$. Then, the expression $F_t = F_0 \circ X_t^{-1}$ admits a unique solution to the initial value problem. Moreover, the following integral formulation

$$F(t,x) = F_0(X_0(t,x)) + \int_0^t f(s,X_s(t,x)) \, ds \quad \text{for every } t \in [0,T] \quad (1.18)$$

solves the Cauchy problem (1.16).

Proof. Given a reference time $s \in [0,T]$, $\mu_0^s \in \mathcal{P}(\mathbb{R}^d)$, and a Borel vector field $v : (t,x) \in [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying the uniform bound condition

$$\int_0^T \left( \sup_{\Omega} |v_t| + \text{Lip}(v_t, \Omega) \right) \, dt < +\infty \quad \text{for every compact set } \Omega \subset \mathbb{R}^d. \quad (1.19)$$

Then, a flow associated to $v_t$, is a map $X(t,s,x) : [0,T] \times [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ such that $t \mapsto X_t(s,x)$ is absolutely continuous in $[0,T]$ and, for every $x \in \mathbb{R}^d$ and $s \in [0,T]$ it is an integral solution of the characteristic system of the DE

$$\partial_t X_t(s,x) = v_t(X_t(s,x)) \quad \text{under } X_s(s,x) = x \quad (1.20)$$

for $\mu_0^s$–a.e. $x \in \mathbb{R}^d$. Indeed, thanks to Proposition 1.2, we can write $X_t(s,x) = y_x(t)$ to denote the solution of the ODE

$$\dot{y}_x(t) = v_t(y_x(t)), \quad y_x(0) = x \quad (1.21)$$

at time $t$, starting from $x$ at the initial times $s$, which admits a unique maximal solution defined in an interval $I(s,x)$ relatively open in $[0,T]$ and containing $s$ as (relatively) internal point. Moreover, if $t \mapsto |X_t(s,x)|$ is bounded in the interior of $I(s,x)$, then $I(s,x) = [0,T]$. Furthermore, if $v$ satisfies the global bounds analogous to (1.19)

$$S := \int_0^T \left( \sup_{\mathbb{R}^d} |v_t| + \text{Lip}(v_t, \mathbb{R}^d) \right) \, dt < +\infty, \quad (1.22)$$
then the flow map \( X \) satisfies the estimates 
\[
\int_0^T \sup_{x \in \mathbb{R}^d} |\partial_t X_t(s, x)| \, dt \leq S \quad \text{and} \quad \sup_{s,t \in [0,T]} \text{Lip} \left( X_t(s, \cdot), \mathbb{R}^d \right) \leq e^S .
\]
In particular, when \( s = 0 \), we will speak of \textit{forward flows}, and we will often omit to indicate the explicit occurrence of \( s \) in \( X_t(0, x) \), writing either \( X_t(x) \) or \( X(t, x) \); analogously, the case \( s = T \) corresponds to \textit{backward flows} (we refer the reader to observations in Section 1.2). As a consequence of the uniqueness of the flow \( X \), the following \textit{semigroup property} holds:

\[
X_t(\tau, x) = X_t(s, X_s(\tau, x)) \quad \text{for every} \quad s, t, \tau \in [0, T].
\]  
(1.23)

Then, \( X_t(\cdot, \cdot) \) are themselves solutions of (1.17): to see this, it suffices to differentiate the semigroup identity (1.23) with respect to \( s \) to obtain, after the change of variables \( x = X_s(\tau, x) \), that the flow \( X_t(s, x) \) satisfies the system of transport equations

\[
\partial_s X_t(s, x) + (v_s(x) \cdot \nabla x) X_t(s, x) = 0 \quad \text{in} \quad [0, T] \times \mathbb{R}^d .
\]  
(1.24)

This solution is also unique with a smooth initial data \( F_0 \), since it is known that any solution of the transport equation has to be constant along characteristics. In particular, as \( X(t, x) = X_t(0, x) = [X_0(t, \cdot)]^{-1}(x) \), we get (1.18): The (formal) proof is based on the simple observation that, we define \( F_t(x) = F_0(X_t^{-1}(t, x)) \) for every \( t \in (0, T) \), then one can easily deduce that it is the only solution of the IVP (1.17). Indeed, we can write \( F_t(X_t(x)) = F_0(x) \) and differentiate in time, thus getting the equation (1.17). Thus, the solution \( F \) is constant on the characteristic lines of \( v \). Consequently, for \( X_s(t, X_t(0, x)) = X_s(0, x) \), we conclude that, for any initial data \( F_0 \), the solution of the transport equation (1.16) is given by the explicit expression (1.18). Here it consists in rewriting the equation as

\[
\frac{d}{dt} [F(t, X_t(0, x))] = f(t, X_t(0, x)), \quad F(0, \cdot) = F_0 ,
\]

which can be integrated directly as

\[
F(t, X_t(0, x)) = F_0(x) - \int_0^t f(s, X_s(0, x)) \, ds .
\]
Therefore, we can deduce that the unique solution of (1.16) is given by

\[ F(t, x) = F_0(X_0(t, x)) - \int_0^t f(s, X_s \circ X_t^{-1}(x)) \, ds. \]

Since \( x \) (and then \( X_t(0, x) \)) is arbitrary we conclude that (1.18) is fulfilled.

As already observed above, the transport equation (1.17) is similar to the continuity equation (1.6), but it is not in conservative form, and the mass of \( F_0 \) is not conserved in time. On the contrary, \( L^\infty \) bounds on \( F_0 \) are conserved, since the values of \( F_t \) are the same as those of \( F_0 \).

**Remark 1.17 (Weak solution and existence)** A locally bounded function \( F : (0, T) \times \mathbb{R}^d \to \mathbb{R} \) is said to be a (weak) solution of (1.17) if the identity

\[ \int_0^T \int_{\mathbb{R}^d} F(\partial_t \xi + \nabla \cdot (v_t \xi)) \, dx \, dt = - \int_{\mathbb{R}^d} \xi(0, x) F_0(x) \, dx \quad (1.25) \]

holds for every test function \( \xi \in C^\infty_c([0, T[ \times \mathbb{R}^d) \). As soon as a vector field \( v_t \) stays Lipschitz continuous, a weak formulation (1.25) is valid for arbitrary \( F_0 \in L^1(\mathbb{R}^d) \), since we can take a sequence of smooth approximating functions \( F^n_0 \to F_0 \), write \( \int_0^T \int_{\mathbb{R}^d} F^n_0 \circ X_t^{-1}(\partial_t \xi + \xi \nabla \cdot v_t + v_t \cdot \nabla \xi) \, dx \, dt \), use the change of variable \( y = X_t^{-1}(x) \), and pass to the limit.

This weak solution can be equivalently defined by noticing that, if \( F \in L^\infty_{\text{loc}}([0, T[ \times \mathbb{R}^d) \), then \( \partial_t F \) has a meaning as a distribution. Furthermore, if we assume that \( \nabla \cdot v \in L^1_{\text{loc}}([0, T[ \times \mathbb{R}^d) \), then the product \( v \cdot \nabla F \) can be defined as a distribution via the equality

\[ \langle \xi, v \cdot \nabla F \rangle = - \langle \xi, F \nabla \cdot v \rangle - \langle \nabla \xi, v F \rangle \]

for every \( \xi \in C^\infty_c([0, T[ \times \mathbb{R}^d) \). This enables us to give directly a distributional meaning to the continuity equation (1.6). The initial data of the Cauchy problem (1.6) can be recovered using Remark 1.9 of weak continuity in time.

We now try to identify the solutions of the Eulerian problem (PDE) (1.6) and we want to connect them to the solutions of the Lagrangian problem (ODE) (1.1) via the flow of the vector field \( v_t \). We recall that the measure denoted by \( X_{t\#} \mu_0 \) is defined through \( X_{t\#} \mu_0(B) := \mu_0(X_t^{-1}(B)) \) for every Borel set \( B \subset \mathbb{R}^d \) and is called the image measure or pushforward of the initial measure \( \mu_0 \) through the map \( X_t : \mathbb{R}^d \to \mathbb{R}^d \) (see Definition 3.2).
**Theorem 1.18** Assume that the vector field \( \mathbf{v} \) is \( C^1 \). Then, for any initial data \( \mu_0 \in \mathcal{P}(\mathbb{R}^d) \), the only measure–valued solution of the continuity equation (1.6) in \( C([0,T];\mathcal{P}(\mathbb{R}^d)) \) is the curve \( (0,T) \ni t \mapsto \mu_t := X_{t\#}\mu_0 \in \mathcal{P}(\mathbb{R}^d) \), where \( \mathcal{P}(\mathbb{R}^d) \) is equipped with the weak topology.

**Proof.** First, it is clear that \( (0,T) \ni t \mapsto \mu_t = X_{t\#}\mu_0 \in \mathcal{P}(\mathbb{R}^d) \) satisfies the continuity requirement (1.8). Indeed, the pushforward definition \( \mu_t = X_{t\#}\mu_0 \) (see Definition 3.2) implies, for every \( \xi \in C^1_c(\mathbb{R}^d) \), the map \( t \mapsto \langle \xi, \mu_t \rangle = \int_{\mathbb{R}^d} \xi(X_t(x)) \, d\mu_0(x) \) belongs to \( C^1([0,T]) \), since \( X \) is \( C^1 \) with respect to time veritable. In fact, the continuity of \( \mu_t \) follows easily since \( \lim_{s \to t} X_s(x) = X_t(x) \) for \( \mu_0 \)-a.e. \( x \in \mathbb{R}^d \): thus for every continuous and bounded function \( \xi : \mathbb{R}^d \to \mathbb{R} \) the dominated convergence theorem yields

\[
\lim_{s \to t} \langle \xi, \mu_s \rangle = \lim_{s \to t} \int_{\mathbb{R}^d} \xi(X_s(x)) \, d\mu_0(x) = \int_{\mathbb{R}^d} \xi(X_t(x)) \, d\mu_0(x) = \langle \xi, \mu_t \rangle.
\]

More precisely, since \( \xi \) is compactly supported and \( X_t^{-1} \) is continuous, the function \( \xi \circ X_t \) is supported in a compact set (uniform for \( t \in [0,T] \)); moreover it is Lipschitz and, for almost every \( t, x, \partial_t (\xi \circ X_t) = \langle \nabla_y \xi \circ X_t, \dot{X}_t \rangle = \langle \nabla_y \xi \circ X_t, \mathbf{v}_t \circ X_t \rangle \). Then, for \( h > 0 \) we can write by using the definition of pushforward that \( \langle (\xi, \mu_{t+h}) - (\xi, \mu_t) \rangle / h = \langle (\xi \circ X_{t+h} - \xi \circ X_t) / h, \mu_0 \rangle \).

Then, the integral of \( (\xi \circ X_{t+h} - \xi \circ X_t) / h \) is uniformly bounded on \( [0,T-h] \times \mathbb{R}^d \), and for almost every \( t \) it converges to \( \langle \nabla \xi \circ X_t \rangle \cdot \mathbf{v}_t \) as \( h \to 0 \), for almost every \( x \). By Lebesgue’s dominated convergence theorem, we deduce that \( t \mapsto \int_{\mathbb{R}^d} \xi(x) \, d\mu_t \) is differentiable for almost every \( t \), and (1.11) holds; that is, \( \frac{d}{dt} \langle \xi, \mu_t \rangle = \langle \nabla \xi, \mu_t \mathbf{v}_t \rangle \). The continuity of \( X_{t\#}\mu_0 \), as a function of \( t \), can be directly applied by Proposition 1.8.

Let us then check that the curve \( (0,T) \ni t \mapsto \mu_t \) satisfies the continuity equation (1.6). Let \( \xi \) be a test function as it is given in Definition 1.7 above, then using the definition \( \mu_t := X_{t\#}\mu_0 \) again, Fubini’s theorem and \( \xi(T,\cdot) = 0 \), we have

\[
\int_0^T \int_{\mathbb{R}^d} \left( \partial_t \xi + \mathbf{v}_t \cdot \nabla \xi \right) \, d\mu_t \, dt = \int_{\mathbb{R}^d} \int_0^T \left( \partial_t \xi(t,X_t(x)) + \mathbf{v}_t(X_t(x)) \cdot \nabla \xi(t,X_t(x)) \right) \, dt \, d\mu_0(x)
= \int_{\mathbb{R}^d} \int_0^T \frac{d}{dt} \left( \xi(t,X_t(x)) \right) \, d\mu_0(x) = - \int_{\mathbb{R}^d} \xi(0,x) \, d\mu_0(x).
\]

So that \( \mu_t := X_{t\#}\mu_0 \) is a measure–valued solution of (1.6).

Now, we turn to uniqueness. By linearity, it is enough to show that if a time–dependent measure \( \mu_t \) (satisfying the regularity conditions of theorem) solves (1.6), then for all \( t \in \]
(0, T), the initial data is \( \mu_0 \) implies the only solution is \( \mu_t = 0 \). We assume that \( t \mapsto \mu^1_t \) and \( t \mapsto \mu^2_t \) are two solutions of (1.6) with \( \mu^0_0 = \mu^2_0 = \mu_0 \) and set \( \mu_t := \mu^1_t - \mu^2_t \), then for every test function \( \xi \in C^1_c([0, T] \times \mathbb{R}^d) \), one has
\[
\int_0^T \int_{\mathbb{R}^d} \left[ \partial_t \xi(t, x) + \mathbf{v}_t(x) \cdot \nabla \xi(t, x) \right] d\mu_t(x) dt = 0. \tag{1.26}
\]
To prove uniqueness we shall use a duality method. Let us consider the backward linear transport equation with a source term \( \zeta \in C^1_c([0, T] \times \mathbb{R}^d) \) on the right hand side:
\[
\partial_t \xi + \mathbf{v}_t \cdot \nabla \xi = \zeta \quad \text{on } (0, T) \times \mathbb{R}^d, \quad \xi(T, \cdot) = \xi_T, \tag{1.27}
\]
where \( \xi_T \) is an arbitrary function in \( C^1_c(\mathbb{R}^d) \). Then, by an argument similar to the one above, we deduce that \( t \mapsto \langle \xi_t, \mu_t \rangle \) is Lipschitz and satisfies \( \frac{d}{dt} \langle \xi_t, \mu_t \rangle = \langle \partial_t \xi_t, \mu_t \rangle + \langle \xi_t, \partial_t \mu_t \rangle = \langle -\mathbf{v}_t \cdot \nabla \xi + \zeta, \mu_t \rangle + \langle \xi_t, \nabla \cdot (\mu_t \mathbf{v}_t) \rangle = \langle \zeta, \mu_t \rangle \) for almost every \( t \). Therefore, we can use \( \zeta \) as a test function in (1.26), which gives that \( \int_0^T \int_{\mathbb{R}^d} \zeta(t, x) d\mu_t(x) dt = 0 \). Since \( \zeta \) is arbitrary function this gives \( \mu_t = 0 \) for almost every \( t \) and we conclude by continuity that \( \mu^1_t(x) = \mu^2_t(x) \) for \( \mathcal{L}^2 \)-a.e. \( (t, x) \in (0, T) \times \mathbb{R}^d \), which implies uniqueness for the solution of the Cauchy problem (1.6).

It remains to construct a solution to (1.27). To solve this equation, we use the classical method of backward characteristics (e.g., \( \mathbf{v} \) satisfies the condition (1.22)), so that solutions are always defined in \( (0, T) \). This means that, thanks to Proposition 1.16, the equation (1.27) has the unique solution
\[
\xi(t, x) = \xi_T(\mathbf{X}_t(x)) - \int_t^T \zeta(s, \mathbf{X}_s(t, x)) ds. \tag{1.28}
\]
Since \( \mathbf{X}_t \) is a locally Lipschitz family and the function \( \zeta \) is compactly supported in space uniformly in time \( t \in (0, T) \), the formula (1.28) does define a Lipschitz function with compact support, satisfying (1.27) almost everywhere.

By “backward transport equation” we mean the following: the transport equation which consists of solving (1.27) for \( t < T \) say, with “arbitrary” data given for \( t = T \) has solutions which do not depend continuously on their initial data. Observe that the characteristics
method for backward first order linear PDE’s provides a useful representation formula for the classical solutions of the backward transport equation (1.27) (formally adjoint to the continuity equation (1.6), in some sense — see Remark 8.1.5 of [17]). Therefore, the explicit expression (1.28) is easily checked with a direct computation, observing that the flow $X_t(s, x)$ of the velocity $v$ starting at time $s \in (0, T)$ satisfies the semigroup property, as a consequence of the uniqueness of the flow (see the proof of Proposition 1.16 above).

**Corollary 1.19** We conclude that, in the sufficiently smooth setting when the vector field $v \in L^1([0, T]; W^{1,\infty}(\mathbb{R}^d; \mathbb{R}^d))$, the solution of (1.6) for any initial data $\mu_0$–a.e. $x \in \mathbb{R}^d$ is

$$\mu_t = X_t\#\mu_0 \quad \text{i.e.} \quad \int_{\mathbb{R}^d} \xi \, d\mu_t = \int_{\mathbb{R}^d} \xi(X_t(x)) \, d\mu_0(x) \quad \text{for every } \xi \in C_c(\mathbb{R}^d), \quad (1.29)$$

and the continuity equation is intended in the usual distributional sense. Furthermore, up to a redefinition of $\mu_t$ in a negligible set of times, the map $t \mapsto \mu_t$ is weakly*–continuous with values in $\mathcal{P}(\mathbb{R}^d)$ (thanks to Remark 1.9). This also gives a sense to the initial data $\mu(0, \cdot) = \mu_0$.

This result has been proved by Ambrosio, Gigli and Savaré [17] (also, see [87]) within minimal regularity assumptions on the velocity field. They used characteristics to prove the existence and representation formulas of the solution of (1.6), under suitable assumption on the velocity $v$ satisfying the globally bound assumption (1.22) made throughout this section.

**Remark 1.20** The presentation of the classical theory outlined above still works (see [6], [12]) under the assumptions of simple local variants $v \in L^1([0, T]; W^{1,\infty}_{\text{loc}}(\mathbb{R}^d; \mathbb{R}^d))$ with a global summability condition

$$\int_0^T \int_{\mathbb{R}^d} \frac{|v_t(x)|}{|x|+1} \, d\mu_t(x) \, dt < +\infty \quad (1.30)$$

for any nonnegative measure $\mu_t$. Under this assumption one can show that for $\mu_0$–a.e. $x$ the maximal solution $X(\cdot, x)$ of the ODE (1.21) starting from $x$ is defined up to $t = T$ and still the representation $\mu_t = X(t, \cdot)\#\mu_0$ holds for $t \in [0, T]$.

We end this subsection by noticing that, in [6], [12], the second growth condition (1.30) on $|v|$ can be replaced by a local, but less intrinsic, condition $|v|/(|x|+1) \in L^1([0, T]; L^\infty(\mathbb{R}^d))$. 
Indeed, due to the growth condition on $v$, we still have pointwise uniqueness of the ODE (1.21) and a uniform local control on the growth of $|X(t, x)|$, therefore we need only to consider a local Lipschitz condition with respect to $x$, integrable with respect to $t$. It is proved by the classical theory of ODE’s that, if $v$ is this regular, then solutions of the ODE (1.21) are unique and stable. We refer the reader to [41], for a brief self-contained exposition of the theory of characteristics for the transport equation and further quantitative estimates in the case of a regular vector field. The fact that the velocity field $v$ is smooth and Lipschitz plays a crucial role through the regularity properties of the flow in the argument above. Understanding what happens precisely in the case of a nonsmooth $v$ has been the subject of an intensive line of research that originated with the fundamental work of DiPerna and Lions [56]. We shall summarize the results of a approach of [56] in the next section.

### 1.4 The Continuity Equation out of the smooth setting

The smooth setting becomes, however, unsatisfactory when one is looking at problems from mathematical physics, mainly systems of conservation laws (again, we recommend [41] for a concise outlook on this matter). The starting point is that the continuity equation, which is a local form of conservation law, can be thought of as a transport equation in which $v$ will depend on the actual density $\rho$. Due to formation of shocks (characteristics crossing), there is no “smooth” theory when it comes to linking the flow equation to the transport equation. As we will see later, in order to obtain existence, uniqueness and stability for the regular Lagrangian flow (see Definition 2.8), it is sufficient to show well-posedness of the continuity equation in the class of bounded weak solutions (see Theorem 2.10). The main strategy of this approach exploits the theory of renormalized solutions for the continuity equation (see Definition 1.23), which is due to DiPerna and Lions [56]. Showing the concept of renormalization property for a given vector field gives uniqueness for the Cauchy problem in the class of bounded weak solutions (see Proposition 1.25), and from this (by now a standard approach) gives existence and uniqueness for a suitable concept of solution to the ODE (1.1), namely the regular Lagrangian flow.
1.4.1 Existence of weak solutions of the continuity equation

Of course, for classical solutions, the identity (1.9) follows from a simple integration by parts. The existence of weak solutions is quite a trivial issue: since the continuity equation is a linear PDE, it is sufficient to regularize the vector field and the initial data, obtaining a sequence of smooth solutions to the approximate problems, and then we pass to the limit to get a solution.

**Theorem 1.21** Let \( v \in L^\infty((0,T) \times \mathbb{R}^d; \mathbb{R}^d) \) with \( \nabla \cdot v \in L^1_{\text{loc}}((0,T) \times \mathbb{R}^d) \) and let \( \mu_0 \in L^\infty(\mathbb{R}^d) \). Then there exists a distributional (weak) solution \( \mu \in L^\infty((0,T) \times \mathbb{R}^d) \) to (1.6).

**Proof.** To prove existence of essentially bounded solutions for the flow problem is easily achieved by smooth approximation even in the case of essentially bounded \( v \) with locally integrable distributional divergence.

Consider a standard family of mollifiers \( \eta^\varepsilon \) and \( \bar{\eta}^\varepsilon \) respectively on \( \mathbb{R}^d \) and \( \mathbb{R}^{1+d} \). Let \( \mu_0^\varepsilon = \mu_0 \ast \eta^\varepsilon \) and \( v^\varepsilon = v \ast \bar{\eta}^\varepsilon \) be the corresponding regularizations of \( \mu_0 \) and \( v \). Then \( \|\mu_0\|_{L^\infty(\mathbb{R}^d)} \) is uniformly bounded. Consider the classical solutions \( \mu^\varepsilon \) of the regularized continuity equation \( \partial_t \mu^\varepsilon + \nabla \cdot (\mu^\varepsilon v^\varepsilon) = 0 \) with \( \mu^\varepsilon(0,\cdot) = \mu_0^\varepsilon \). This solution exists and is unique for every \( \varepsilon > 0 \), since the vector field \( v^\varepsilon \) is smooth we can apply the classical Cauchy–Lipschitz theorem to solve (1.3).

From the explicit formula (1.29), for the solution of the continuity equation with smooth vector field, we immediately deduce that the sequence \( (\mu^\varepsilon) \) is uniformly bounded in \( L^\infty((0,T) \times \mathbb{R}^d) \). Indeed, by Maximum Principle Lemma (see for example, Lemma 1.2 of [51]): for every \( t \), we have \( \sup_{x \in \mathbb{R}^d} \mu(t,x) \leq \sup_{x \in \mathbb{R}^d} \mu_0(x) \) and \( \inf_{x \in \mathbb{R}^d} \mu(t,x) \geq \inf_{x \in \mathbb{R}^d} \mu_0(x) \) implies \( \|\mu(t,\cdot)\|_{L^\infty(\mathbb{R}^d)} \leq \|\mu_0\|_{L^\infty(\mathbb{R}^d)} \). Thus, we can find a subsequence \( (\mu^{\varepsilon_k}) \), which converges weakly in \( L^\infty((0,T) \times \mathbb{R}^d) \) to a function \( \mu \in L^\infty(w^\ast((0,T) \times \mathbb{R}^d)) \). Recalling the weak formulation (1.9) and replacing \( \mu, v \) and \( \mu_0 \) with \( \mu^\varepsilon, v^\varepsilon \) and \( \mu_0^\varepsilon \), it is immediate to deduce that \( \mu \) is a solution of (1.6): since \( v^\varepsilon \to v \) and \( \mu^\varepsilon \to \mu \) locally strongly in \( L^1_{\text{loc}} \), we can pass into the limit in such identities to achieve (1.9) for \( \mu, v \) and \( \mu_0 \). \( \square \)
1.4.2 Vector fields with Sobolev spatial regularity

Since the existence of distributional solution has already been shown for the continuity equation, we now want to face the uniqueness issue. Of course, the next relevant questions are whether such distributional solutions are unique, and stable. Under the general assumptions above the answer is negative, as it is for instance showed by the elegant example of [54], which yields an explicit counterexample to the uniqueness (see Chapter 5 of [41] for more details on such examples and related constructs). However, DiPerna and Lions in [56] proved stability and uniqueness when a bounded vector field $v_t$ belonging to $W^{1,p}$ with bounded spatial divergence.

**Theorem 1.22 (DiPerna–Lions)** Let $v \in L^\infty((0,T) \times \mathbb{R}^d; \mathbb{R}^d) \cap L^1([0,T]; W^{1,p}(\mathbb{R}^d; \mathbb{R}^d))$ be a vector field such that $\nabla \cdot v \in L^\infty((0,T) \times \mathbb{R}^d)$. Then, for every $\mu_0 \in L^\infty(\mathbb{R}^d)$ there exists a unique and stable distributional solution of (1.6). By stability we mean the following: let $v_h$ and $\mu_h^0$ be two smooth approximating sequences converging strongly in $L^1_{\text{loc}}$ to $v$ and $\mu_0$ respectively, with $\|\mu_h^0\|_\infty$ uniformly bounded; then the solutions $\mu_h^0$ of the corresponding continuity equations converge strongly in $L^1_{\text{loc}}$ to the solution $\mu$ of (1.6).

**Proof.** The proof of Theorem 1.22 by DiPerna and Lions consists of two parts, the first one, which is “soft” can be stated as Proposition 1.25 below: If $v \in L^\infty$ has the renormalization property and its divergence is bounded, then the uniqueness and stability properties of Theorem 1.22 hold. The second one, which is the “hard” part of the proof, states essentially that $W^{1,p}$ fields have the renormalization property. We postpone the “hard part” to the next step and come first to Proposition 1.25.

**Step 1 (Renormalized solutions).** In order to understand their proof, we first go back to classical solutions $\mu$ of (1.6), and we observe that, whenever $\phi : \mathbb{R}^d \to \mathbb{R}^d$ is a $C^1$ function, $\phi \circ \mu$ solves a “slightly” modified equation

$$\partial_t \phi(\mu) + \nabla \cdot (v \phi(\mu)) = (\nabla \cdot v)[\phi(\mu) - \mu \phi'(\mu)], \quad \text{under} \quad \phi(\mu(0,\cdot)) = \phi \circ \mu_0. \quad (1.31)$$

This can be seen, for instance, using the chain rule for differentiable functions, i.e. $\partial_t \phi(\mu) + \nabla \cdot (v \phi(\mu)) = \phi'(\mu)[\partial_t \mu + v \cdot \nabla \mu] + \phi(\mu) \nabla \cdot v$. Only if $\phi$ is linear or $\nabla \cdot v \equiv 0$, then we
get the renormalized solution $\phi \circ \mu$ solves the continuity equation $\partial_t \phi(\mu) + \nabla \cdot (\mathbf{v} \phi(\mu)) = 0$. Otherwise, we dealing with the continuity equations of the form $\partial_t \mu + \mathbf{v} \cdot \nabla \mu + \mu \nabla \cdot \mathbf{v} = 0$. One also can observe that, since the transport equations (1.17) is linear and its solution $F$ must be constant along the trajectories (1.3), so the renormalized solution of $\partial_t \phi(F) + \mathbf{v} \cdot \nabla \phi(F) = 0$ must be $\phi \circ F$. Motivated by these observations, we introduce the following terminology and define the notion of renormalized solution corresponds to the idea of “solution which satisfies the desired chain rule”.

**Definition 1.23 (Renormalized solution)** Let $\mathbf{v} \in L^1_{\text{loc}}((0, T) \times \mathbb{R}^d; \mathbb{R}^d)$ be a vector field such that $\nabla \cdot \mathbf{v} \in L^1_{\text{loc}}((0, T) \times \mathbb{R}^d)$. A distributional solution $\mu \in L^\infty((0, T) \times \mathbb{R}^d)$ to (1.6) relative to $\mathbf{v}$ with initial data $\mu_0 \in L^\infty(\mathbb{R}^d)$ is said to be a renormalized solution if for every function $\phi \in C^1(\mathbb{R})$, the identity (1.31) holds in the sense of distributions in $[0, T] \times \mathbb{R}^d$. Analogously, for every $\xi \in C^\infty_c([0, T] \times \mathbb{R}^d)$

$$
\int_0^T \int_{\mathbb{R}^d} (\partial_t \xi + \mathbf{v} \cdot \nabla \xi) \phi(\mu) \, dx \, dt + \int_0^T \int_{\mathbb{R}^d} \xi(\nabla \cdot \mathbf{v} \phi(\mu) - \mu \phi'(\mu)) \, dx \, dt = -\int_{\mathbb{R}^d} \xi(0, x) \phi(\mu_0) \, dx.
$$

The assumption $\phi \in C^1(\mathbb{R})$ is exploited to give a distributional meaning to the equality (1.31), which becomes locally integrable despite the lack of integrability of $\mu$. We refer the reader to [40] for renormalized solutions to the continuity equation in the case of an integrable damping term.

This definition was first introduced in [56] to the nonhomogeneous linear PDE (1.16), namely the transport equations with source term $f \in L^1((0, T); L^\infty_{\text{loc}}(\mathbb{R}^d))$. It basically proved that whenever $F$ is a weak solution for the transport equation, then so is $\phi \circ F$ for smooth $F$. Obviously, classical solutions, if they existed, would satisfy that. In particular, choosing $f = -\nabla \cdot \mathbf{v} \, F$, we are able to translate all the well-posedness results for the transport equation into well-posedness results for the continuity equation with minor modifications in a proof (see, for example [12], [41] for a more details on such problems). However, when the vector field is not smooth, we cannot expect any regularity of the solutions, so that (1.31) is a nontrivial request when made for all bounded distributional solutions.

**Remark 1.24** When dealing with the continuity equation (1.6) we can consider the renor-
malization function $\phi(z) = \sqrt{1 + (z^+) - 1} \in C^1(\mathbb{R})$. Integrating formally over $\mathbb{R}^d$ we obtain
\[
\frac{d}{dt} \int_{\mathbb{R}^d} \phi(\mu(t, x)) \, dx = \int_{\mathbb{R}^d} \nabla \cdot v(t, x) \left[ \phi(\mu(t, x)) - \mu(t, x) \phi'(\mu(t, x)) \right] \, dx
\]

Since for this choice of $\phi$ we have $\phi(z) - z\phi'(z) \leq 0$, it turns out that it is enough to have a control on the negative part of the divergence: we only need to assume $[\nabla \cdot v]^- \in L^1([0, T]; L^\infty(\mathbb{R}^d))$.

Notice that when $\phi(z) = z^2$ the identity
\[
\partial_t(\mu^2) + \nabla (\mu^2 v) = -\mu^2 \nabla \cdot v
\] (1.32)
is trivially satisfied if $\mu$ is smooth, by an immediate application of the chain rule formula. The property of renormalization, if satisfied by all bounded weak solutions, can be transferred into a property of the vector field itself. We say that a vector field $v$ has the renormalization property if every bounded weak solution $\mu$ of the continuity equation (1.6) with vector field $v$ is a renormalized solution. In fact, the renormalization property to the transport equation (1.17) asserts that nonlinear compositions of the solution are again solutions, or alternatively that the chain rule holds in this weak context. The importance of this notion comes from the fact that, if the renormalization property of a vector field $v$ holds, the continuity equation (1.6) is well-posed. The idea, at least at a formal level, is quite simple: we simply integrate (1.32) over $\mathbb{R}^d$ for every fixed time $t$, and assuming a sufficiently fast decay at infinity we obtain
\[
\frac{d}{dt} \int_{\mathbb{R}^d} \mu^2(t, x) \, dx = -\int_{\mathbb{R}^d} \mu^2(t, x) \nabla \cdot v(t, x) \, dx.
\]
Assuming an $L^\infty$ control on the divergence of $v$ we can deduce
\[
\frac{d}{dt} \int_{\mathbb{R}^d} \mu^2(t, x) \, dx \leq \|\nabla \cdot v\|_{L^\infty(\mathbb{R}^d)} \int_{\mathbb{R}^d} \mu^2(t, x) \, dx,
\]
thus a simple application of the Gronwall inequality implies that, if $\mu_0 = 0$, we have $\int_{\mathbb{R}^d} \mu^2(0, x) \, dx = 0$, and hence we would conclude $\int_{\mathbb{R}^d} \mu^2(t, x) \, dx = 0$ for every $t$. We sketch a proof of how to make rigorous this formal argument. Fix $T, \tau > 0$ and con-
sider a test function \( \xi \in C^\infty_c((-\infty, T) \times \mathbb{R}^d) \) such that \( \xi \equiv 1 \) on \([0, T - \tau] \times B_{\mathbb{R}^d}(0) \) and \( \partial_t \xi(t, x) \leq -\|v\|_{\infty} |\nabla \xi(t, x)| \) on \([0, T] \times \mathbb{R}^d \). Now let \( \zeta \in C^\infty_c((-\infty, T)) \) be nonnegative and test (1.32) with \( \zeta(t) \xi(t, x) \). Define \( f(t) = \int_{\mathbb{R}^d} \mu^2(t, x) \xi(t, x) \, dx \) and use Fubini’s Theorem to get

\[
- \int_{-\infty}^{T} f(t) \zeta'(t) \, dt = \int_{-\infty}^{T} \int_{\mathbb{R}^d} \zeta(t) \mu^2(t, x) [\partial_t \xi(t, x) + \nabla \cdot \xi(t, x)] \, dx \, dt \\
+ \int_{-\infty}^{T} f(t) \nabla \cdot (v(t, x) \zeta(t)) \, dt
\]

Note that the first integral in the right hand side is nonpositive, whereas the second one can be estimated by \( \|\nabla \cdot v(t, \cdot)\|_{L^\infty(\mathbb{R}^d)} f(T) \zeta(T) \, dt \). We conclude that \( f \) satisfies a “distributional” form of Gronwall’s inequality for \( t \in (-\infty, T] \). This means that we have \( f'(t) \leq \|\nabla \cdot v(t, \cdot)\|_{L^\infty(\mathbb{R}^d)} f(t) \) in the sense of distributions in \((-\infty, T] \). It can be easily seen that this implies \( f \equiv 0 \). Thus \( \mu \equiv 0 \) a.e. on \([0, T] \times B_{\mathbb{R}^d}(0) \), and by the arbitrariness of \( R, T \) and \( \xi(t, x) \) we conclude \( \mu_t = 0 \) for every \( t \in [0, T] \), as desired.

**Proposition 1.25** Let \( v \in L^\infty((0, T) \times \mathbb{R}^d) \) with \( [\nabla \cdot v]^- \in L^1((0, T); L^\infty(\mathbb{R}^d)) \). Extend the vector field \( v \) to negative times by setting \( v \equiv 0 \) for \( t < 0 \). If \( v \) has the renormalization property on \((-\infty, T) \times \mathbb{R}^d \), then bounded solutions of the continuity equation (1.6) are unique and stable.

**Proof.** The proof of uniqueness immediately follows from the discussion in the previous paragraph (also, see for example Theorem 2.3.3 of [41], Proposition 1.6 of [51], or Theorem 27 of [13] for a more general discussion). Being the continuity equation linear, it is enough to prove that the only bounded solution to the Cauchy problem (1.6) with initial data 0 is \( \mu \equiv 0 \).

The only modification in the proof consists in taking into account an additional assumption \( [\nabla \cdot v]^- \in L^1((0, T); L^\infty(\mathbb{R}^d)) \) which is clearly concluded from \( \nabla \cdot v \in L^1((0, T); L^\infty(\mathbb{R}^d)) \).

Arguing as in Theorem 1.21, we easily deduce that, up to subsequences, \( (\mu_h) \) converges in \( L^\infty_{w^*}((0, T) \times \mathbb{R}^d) \) to a distributional solution \( \mu \) of (1.6). However, by the uniqueness part of the theorem, we know that this solution is unique, and then the whole sequence converges to \( \mu \). Since \( v_h \) and \( \mu_h \) are both smooth, \( \mu^2_h \) solves the continuity equation with vector field \( v_h \) and initial data \( \mu^2_0(0, \cdot) \). Arguing as before, \( \mu^2_h \) must converge in \( L^\infty_w((0, T) \times \mathbb{R}^d) \) to the unique solution of (1.6) with initial data \( \mu^2_0 \). But by the renormalization property this
solution is \( \mu^2 \). Hence we have shown that \( \mu_h \rightharpoonup \mu \) and \( \mu_h^2 \rightharpoonup \mu^2 \) in \( L^\infty((0,T) \times \mathbb{R}^d) \), and this eventually implies \( \mu_h \to \mu \) strongly in \( L^1_{\text{loc}}((0,T) \times \mathbb{R}^d) \).

\[ \square \]

**Remark 1.26** In the previous theorem the extension to negative times is necessary: it is a way of requiring the strong continuity of the solution at the initial time (see the discussion in Section 2.1 and Theorem 3.1.1 of [41]). Another possibility would be to change the definition of renormalized solution, asking that \( \phi \circ F \) should satisfy the transport equation (1.17) and have initial data \( \phi \circ F_0 \) (this is the approach of [51]).

**Remark 1.27** The renormalization property for the extended vector field in general does not follow from the renormalization property for \( \mathbf{v} \). Moreover, even if \( \mathbf{v} \) has the renormalization property in the sense of Definition 1.23, there could exist a solution \( \mu \) which is renormalized, but such that \( \phi \circ \mu(0,\cdot) \neq \phi \circ \mu_0 \). An explicit example is given by the oscillatory solution constructed by Depauw's counterexample [54] and presented in Section 5.1 of [41]. In this case the vector field \( \mathbf{v} \) has the renormalization property (in \( (0,1) \times \mathbb{R}^2 \)) but the extension of \( \mathbf{v} \) has not the renormalization property in \( (-\infty,1) \times \mathbb{R}^2 \); the Cauchy problem for the continuity equation with initial data \( \mu_0 \equiv 0 \) has more than one solution. This example also shows the sharpness of the assumptions of Theorem 2.6.1: indeed, \( \mathbf{v} \in L^1_{\text{loc}}((0,1);BV_{\text{loc}}(\mathbb{R}^2;\mathbb{R}^2)) \) but \( \mathbf{v} \notin L^1_{\text{loc}}((0,1);BV_{\text{loc}}(\mathbb{R}^2;\mathbb{R}^2)) \), thus the extension of \( \mathbf{v} \) does not belong to \( L^1_{\text{loc}}((\mathbb{R}^2;\mathbb{R}^2)) \).

**Step 2 (The commutator estimate of DiPerna and Lions).** In this step we come to prove the “hard part” for Theorem 1.22. We first prove a milder conclusion, neglecting the initial conditions, which will be adjusted later.

**Proposition 1.28** Any \( \mathbf{v} \in L^1([0,T];W^{1,p}(\mathbb{R}^d;\mathbb{R}^d)) \) has the renormalization property, i.e. whenever \( \mu \in L^\infty((0,T) \times \mathbb{R}^d) \) solve the continuity equation (1.6) distributionally on \( (0,T) \times \mathbb{R}^d \), then so does \( \phi \circ \mu \) for every \( \phi \in C^1(\mathbb{R}^d) \).

Note that, assuming the velocity \( \mathbf{v} \) lies in \( L^1([0,T];W^{1,p}(\mathbb{R}^d;\mathbb{R}^d)) \) means that it is not smooth but have some Sobolev differentiability (or Sobolev regularity) with respect to spatial variable \( x \); that is, \( \mathbf{v}, \nabla \cdot \mathbf{v} \in L^1([0,T];L^p(\mathbb{R}^d;\mathbb{R}^d)) \), where \( \nabla \cdot \mathbf{v} \) is the gradient of Jacobian (since \( \mathbf{v} \) is a vector field).
The main tool to show the renormalization property is a regularization scheme due to DiPerna and Lions. Since smooth solutions are renormalized, a natural attempt consists in convolving the continuity equation (1.6) with a regularization kernel \( \eta^\varepsilon \) in \( \mathbb{R}^d \), using the classical chain rule for the regularized solution \( \mu \ast \eta^\varepsilon \) and finally trying to pass to the limit in the equality. In order to do this, let us fix \( \mu \) and \( \mathbf{v} \) as in Proposition (1.28) and consider a standard smooth and even kernel \( \eta \) in \( \mathbb{R}^d \). By a slight abuse of notation we denote by \( \mu \ast \eta^\varepsilon \) the convolution in the \( x \) variable, that is \( \mu_t \ast \eta^\varepsilon (x) = \int_{\mathbb{R}^d} \mu_t(y) \eta^\varepsilon (x - y) \, dy \). Mollify (1.6) to obtain \( 0 = \partial_t \mu_t \ast \eta^\varepsilon + \nabla \cdot (\mathbf{v}_t \mu_t) \ast \eta^\varepsilon \). We define \( \mu^\varepsilon_t = \mu_t \ast \eta^\varepsilon \) and we notice that we can let the convolution act on \( \mu_t \) in \( \nabla (\mathbf{v}_t \mu_t) \) only at the price of an error term, so we can rewrite this identity as

\[
\partial_t \mu^\varepsilon_t + \nabla \cdot (\mathbf{v}_t \mu^\varepsilon_t) = R^\varepsilon + \mu^\varepsilon_t \nabla \cdot \mathbf{v}_t - (\mu_t \nabla \cdot \mathbf{v}_t) \ast \eta^\varepsilon .
\]  

(1.33)

where the error term \( R^\varepsilon \) in the right hand side are simply the **commutators**

\[
R^\varepsilon := \mathbf{v}_t \cdot \nabla (\mu^\varepsilon_t) - (\mathbf{v}_t \cdot \nabla \mu_t) \ast \eta^\varepsilon .
\]  

(1.34)

This term is called commutator because it measures the difference in exchanging the two operations of convolution and of differentiation in the direction of \( \mathbf{v} \). Now notice that the function \( \mu^\varepsilon_t \) is smooth with respect to the spatial variable. Since \( R^\varepsilon \) is a locally summable function, the identity (1.33) implies that \( \partial_t \mu^\varepsilon_t \) is also locally summable. Thus, for every \( \varepsilon > 0 \) fixed, \( \mu^\varepsilon_t \) has Sobolev regularity in space and time, and we can use the chain rule for Sobolev functions (see for instance Section 4.2.2 of [61]) to compute

\[
\partial_t \phi (\mu^\varepsilon_t) + \nabla \cdot (\mathbf{v}_t \phi (\mu^\varepsilon_t)) = \phi'(\mu^\varepsilon_t) [\partial_t \mu^\varepsilon_t + \nabla \cdot (\mathbf{v}_t \mu^\varepsilon_t)] + (\nabla \cdot \mathbf{v}_t) [\phi (\mu^\varepsilon_t) - \mu^\varepsilon_t \phi' (\mu^\varepsilon_t)].
\]

Inserting (1.33) in this identity we get

\[
\partial_t \phi (\mu^\varepsilon_t) + \nabla \cdot (\mathbf{v}_t \phi (\mu^\varepsilon_t)) = \phi'(\mu^\varepsilon_t) [R^\varepsilon + \mu^\varepsilon_t \nabla \cdot \mathbf{v}_t - (\mu_t \nabla \cdot \mathbf{v}_t) \ast \eta^\varepsilon ] + (\nabla \cdot \mathbf{v}_t) [\phi (\mu^\varepsilon_t) - \mu^\varepsilon_t \phi' (\mu^\varepsilon_t)].
\]  

(1.35)

Now, the left hand side of (1.35) converges distributionally to the left hand side of \( \partial_t \phi (\mu) + \)
\[ \nabla \cdot (v_t \phi(\mu)) = (\nabla \cdot v_t)[\phi(\mu_t) - \mu_t \phi'(\mu_t)]. \]

Recall that \( \|\phi'(\mu_t^\varepsilon)\|_\infty \) and \( \|\mu_t^\varepsilon\|_\infty \) are uniformly bounded, whereas

\[ \mu_t^\varepsilon \nabla \cdot v_t - (\mu_t \nabla \cdot v_t) * \eta^\varepsilon \to 0 \text{ strongly in } L^1_{\text{loc}}. \]

Therefore, in order to prove Proposition 1.28 we just need to show that \( \phi'(\mu_t^\varepsilon) R^\varepsilon \) converges to zero. Notice that \( R^\varepsilon \) converges to zero in the sense of distributions, without any regularity assumption on \( v \). The sequence \( \phi'(\mu_t^\varepsilon) \) is equibounded in \( L^\infty \) and converges \( L^d \)–a.e. for every \( t \in (0, T) \), but this does not allow to deduce the convergence of the product \( \phi'(\mu_t^\varepsilon) R^\varepsilon \). It turns out that it is necessary to study the behaviour of the commutator error term \( R^\varepsilon \). A first attempt would be to show the strong convergence to zero of the commutator \( R^\varepsilon \). This is precisely the result that DiPerna and Lions obtained in [56] under a Sobolev regularity assumption. This is implied by the following lemma.

**Lemma 1.29 (Commutator estimate)** Let \( v \in L^1([0, T]; W^{1,p}(\mathbb{R}^d; \mathbb{R}^d)), \mu \in L^\infty((0, T) \times \mathbb{R}^d) \) and \( R^\varepsilon \) as in (1.34). Then, \( R^\varepsilon \to 0 \) strongly in \( L^1_{\text{loc}}((0, T) \times \mathbb{R}^d) \).

**Proof.** Without loss of generality we assume that the kernel \( \eta \) is supported in \( B_1(0) \). First playing with the definitions of \( v \cdot \nabla \mu \) and of the convolution of a distribution and a smooth function we have the elementary identity of the commutator

\[ R^\varepsilon = \sum_i v_i^\varepsilon(\mu_t \partial_x \eta^\varepsilon) - \sum_i (\mu_t v_i^\varepsilon) \partial_x \eta^\varepsilon + (\mu_t \nabla \cdot v_t) \eta^\varepsilon \]

and we expand the convolutions to obtain the following explicit expression

\[ R^\varepsilon(t, x) = \int_{B_1(0)} \mu_t(x)(v_t(x) - v_t(y)) \cdot \nabla \eta^\varepsilon(x - y) dy + (\mu_t \nabla \cdot v_t) \eta^\varepsilon. \]

Since \( \nabla \eta^\varepsilon(z) = \varepsilon^{-d-1} \nabla \eta(z/\varepsilon) \), we perform the change of variables \( z = (x - y)/\varepsilon \) to get

\[ R^\varepsilon(t, x) = \int_{B_1(0)} \mu_t(x - \varepsilon z) \frac{v_i(x - \varepsilon z) - v_i(x)}{\varepsilon} \cdot \nabla \eta(z) dz + (\mu_t \nabla \cdot v_t) \eta^\varepsilon. \]  

(1.36)

Next, fix a compact set \( \Omega \subset \mathbb{R}^d \). By standard properties of Sobolev functions (see for instance Section 5.8.2 of [59] or Theorem 2.1.6 of [120]), the directional difference quotients
of \( \mathbf{v} \) in the direction of \( z \) at the point \( x \)

\[
D_z \mathbf{v}(t, \cdot) := \frac{\mathbf{v}_t(x) - \mathbf{v}_t(x - \varepsilon z)}{\varepsilon}
\]

(1.37)

are bounded in \( L^p(\Omega) \) independently of \( z \in B_1(0) \) and \( 0 < \varepsilon < 1 \). We now let \( \varepsilon \) goes to zero. For each fixed \( z \), (1.37) converges strongly in \( L^p(\Omega) \) to \( \partial_z \mathbf{v}_t \). The functions \( \mu_t(x - \varepsilon z) \) are instead uniformly bounded in \( L^\infty(\mathbb{R}^d) \), and, by the \( L^1 \)–continuity of the translation, they converge strongly in \( L^1(\Omega) \) to \( \mu_t \). Therefore, we conclude that \( R^\varepsilon \) converges strongly in \( L^1_{\text{loc}}((0, T) \times \mathbb{R}^d) \) to

\[
R^0(t, x) = \mu_t(x) \int_{\mathbb{R}^d} \partial_z \mathbf{v}_t \cdot \nabla \eta(z) \, dz + (\mu_t \nabla \cdot \mathbf{v}_t)(x)
\]

\[
= \mu_t(x) \sum_{i,j} \partial_i \mathbf{v}_t^j \int_{\mathbb{R}^d} z_i \partial_j \eta(z) \, dz + \mu_t(x) \nabla \cdot \mathbf{v}_t(x)
\]

Integrating by parts we have \( \int_{\mathbb{R}^d} z_i \partial_j \eta = -\delta_{ij} \). So \( R^0 = 0 \), which completes the proof. \( \Box \)

**Remark 1.30** In order to show convergence to zero in a sufficiently strong sense of the commutator \( R^\varepsilon \) it is necessary (at least at a first level) to have some control on the behaviour of the difference quotients of the vector field \( \mathbf{v} \). Here the weak differentiability assumptions on \( \mathbf{v} \) come into play: the renormalization property can be proved, for instance, if \( \mathbf{v} \) has spatial Sobolev regularity (DiPerna and Lions [56]) or BV regularity (Ambrosio [6]) with respect to the spatial variable, and under boundedness assumptions on the divergence of the vector field. For a survey of some other renormalization results see for instance Chapter 2 of [41].

**Step 3 (The initial condition).** In order to prove Theorem 1.22 we still need to show that \( \phi \circ \mu_t \) takes the initial condition \( \phi(\mu)(0, \cdot) = \phi(\mu_0) \). This is achieved with a small trick. Consider \( \mathbf{v} \) and \( \mu \) as in Theorem 1.22 and extend both of them to negative times by setting \( \mathbf{v}(t, x) = 0 \) and \( \mu(t, x) = \mu_0(x) \) for \( t < 0 \). It is then immediate to check that \( \partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \) distributionally on the time–spac \( (-\infty, T) \times \mathbb{R}^d \). On the other hand the proof of Proposition 1.28 remains valid if we replace \( (0, T) \) with \( (-\infty, T) \) (actually the proof is the same on any open set \( \mathcal{O} \subset \mathbb{R} \times \mathbb{R}^d \)). Therefore \( \partial_t \phi(\mu) + \nabla \cdot (\mathbf{v} \phi(\mu)) = (\nabla \cdot \mathbf{v})[\phi(\mu) - \mu \phi'(\mu)] \) distributionally on \( (-\infty, T) \times \mathbb{R}^d \). We test this equation with a \( \xi \in C^\infty_c((-\infty, T) \times \mathbb{R}^d) \),
recalling that $\phi(\mu)(0, x) = \phi(\mu_0(x))$ and $v(t, x) = 0$ for $t < 0$. We then conclude

$$
\int_0^T \int_{\mathbb{R}^d} \left( [\partial_t \xi + v \cdot \nabla \xi] \phi(\mu) + \xi [\nabla \cdot v(\phi(\mu) - \mu \phi'(\mu))] \right) dx \, dt = - \int_{\mathbb{R}^d} \phi(\mu_0(x)) \int_{-\infty}^0 \partial_t \xi(t, x) \, dt \, dx.
$$

On the other hand, since $\xi$ is smooth, we can integrate by parts in $t$ in the right hand side of the above identity in order to get $- \int_{\mathbb{R}^d} \phi(\mu_0(x))\xi(0, x) \, dx$. This concludes the whole proof of Theorem 1.22.

\[\square\]

### 1.4.3 Vector fields with BV spatial regularity

In this subsection, we present the generalized result of a well-posedness theory in the class of vector fields $v$ with bounded variation. As we mentioned earlier that, the same results of Theorem 1.22, in the bounded variation case, have been already obtained by Ambrosio [6] with different techniques. The main point, which makes the difference with respect to the Sobolev case, is the lack of strong convergence of the difference quotients (compare with the proof of Lemma 1.29). In fact, Ambrosio developed a uniqueness theory based on the renormalization property of a vector field with $BV$ regularity with respect to the spatial variable. More precisely, existence and uniqueness of bounded weak solutions to the initial value problem for (1.6) was proved when $v \in L^1(0, T; BV(\mathbb{R}^d, \mathbb{R}^d))$ and $\nabla \cdot v \in L^1(0, T; L^\infty(\mathbb{R}^d))$. The general strategy used in [6] is similar to the one used in the previous subsection, however the proof of the renormalization property is more difficult than in the proof of Theorem 1.22 and involves convolutions with anisotropic kernels and Albertis rank-one theorem [1].

Note that, as in Theorem 1.22, in [6] the assumption $\nabla \cdot v \in L^1(0, T; L^\infty(\mathbb{R}^d))$ is used only in the proof of existence of weak solutions of (1.6), while the assumption $\nabla \cdot v \in L^1((0, T) \times \mathbb{R}^d)$ is sufficient for uniqueness of weak solutions and for the renormalization property. Therefore, one of the possible directions for developing further the theory of (1.6) is to go beyond the assumption of absolute continuity of $\nabla \cdot v$ with respect to Lebesgue measure $\mathcal{L}^d$. As such, Ambrosio extended, in his celebrated paper [6], the renormalization property in Proposition 1.28 to the case of a $BV$ dependence with respect to the spatial variables, but under the assumption that the divergence of $v$ is absolutely continuous.
Theorem 1.31 (Ambrosio) Let \( \mathbf{v} \in L^1_{\text{loc}}((0,T); BV_{\text{loc}}(\mathbb{R}^d, \mathbb{R}^d)) \) such that \( \nabla \cdot \mathbf{v} \ll \mathcal{L}^d \) for \( \mathcal{L}^1 \)-a.e. \( t \in (0,T) \). Then, \( \mathbf{v} \) has the renormalization property. Moreover, for every \( \mu_0 \in L^\infty(\mathbb{R}^d) \), there exists a unique and stable distributional solution \( \mu \in L^\infty((0,T) \times \mathbb{R}^d) \).

The proof involves a splitting of the difference quotients and an anisotropic regularization, based on a local selection of a “bad” direction given by Alberti’s Rank–one theorem [1]. Grippa [41] also gave a reasonably detailed description of the proof of this result, mostly trying to enlighten the main ideas. He presented an approach which is slightly simpler than the one of the original paper [6]. We refer to [5], [6], [7] and [51] for an account of the proof of Ambrosio’s theorem. This theorem has been applied to the study of various nonlinear PDEs: for instance the Keyfitz and Kranzer system ([14] and [8]) and the semigeostrophic equation ([47], [48] and [26]).

The assumption \( \nabla \cdot \mathbf{v} \ll \mathcal{L}^d \) (in Theorem 1.31) is used in the definition of weak solution of (1.6). Therefore, if \( \nabla \cdot \mathbf{v} \) is not absolutely continuous, we still have to impose some additional restriction on the divergence of \( \mathbf{v} \), which would allow us to give a meaning to (1.6). One of such restrictions is the assumption that \( \mathbf{v} \) is nearly incompressible, i.e. there exists a “density” function \( \rho \in L^\infty((0,T) \times \mathbb{R}^d) \) and a constant \( C > 0 \) such that \( 0 < C^{-1} < \rho < C < +\infty \) for \( \mathcal{L}^{d+1} \)-a.e. \( (t,x) \in (0,T) \times \mathbb{R}^d \) and the identity \( \partial_t \rho + \nabla \cdot (\mathbf{v} \rho) = 0 \) holds in the sense of distributions in \( (0,T) \times \mathbb{R}^d \). We refer the reader to [8], [14], [41], [50] for studying a well–posedness theory for vector fields in which the usual assumption of boundedness of the divergence is replaced by a control of the Jacobian (or, by the existence of a solution of the continuity equation, which is bounded away from 0 and +\( \infty \)).
Chapter 2

Theory of regular Lagrangian flows

In the last years much attention has been devoted to the study of flows associated to vector fields that are not smooth (in particular, less than Lipschitz in the space variable, but just in some weak differentiability class). In this chapter, we present results were studied by DiPerna–Lions [56] and Ambrosio [6] in a nonclassical setting about the relationship between the continuity equation (1.6) and the associated first order ODE (1.1), and about uniqueness of the PDE versus uniqueness of the ODE. We shall recall the definition of regular Lagrangian flow and describe some well–posedness results that are available in the high–dimensional case. The authors of [6], [56] (also, see [12], [13], [15], [41], [42], [43], [44], [45], [46] for a good survey on such problems) have addressed the questions of existence, uniqueness and stability for regular Lagrangian flows out of the smooth setting. In Section 2.1, we present the probabilistic point of view. In order to study existence, uniqueness and stability of solutions to the ODE out of the smooth setting, we consider suitable measures in the space of continuous maps, allowing for probabilistic representation of trajectories. Then, in some special situations we are able to show that this probabilistic representation actually does not occur, but still this “probabilistic” interpretation is very useful to understand the underlying techniques and to give an intrinsic characterization of the flow. In Section 2.2, the equivalence between pointwise uniqueness for solutions of the ODE and uniqueness of positive measure–valued solutions of the continuity equation is shown, exploiting the probabilistic representation for solutions of the PDE. Section 2.3.3 is devoted to the theory of regular Lagrangian flows, developed by Ambrosio in [6] (see also [6], [13], [41]) and we
show how the uniqueness of bounded solutions of the continuity equation implies existence and uniqueness of the regular Lagrangian flow. We introduce, heuristically, the concept of “regular Lagrangian flow” and how it will be the good notion of a solution to the ODE in the nonsmooth context. Finally in the last section, we comment briefly on the notion of Lagrangian flow defined by DiPerna and Lions.

2.1 The probabilistic point of view

In order to illustrate the motivations which led us to the results of this section, we assume, for a moment, that the vector field $v$ (satisfying the condition (1.22)) is sufficiently regular, in such a way that, for $\mu_0$–a.e. $x \in \mathbb{R}^d$, the characteristic system (1.20) admits a unique globally defined solution $X_t(x)$ in $[0,T]$. In this case and thanks to Corollary 1.19, it is well known that the solution of the continuity equation (1.6) is representable by the formula (1.29). More precisely, we consider a continuous (in the sense of narrow topology, see Section 5.1 [17]) time–dependent family $\mu_t$, $t \in [0,T]$, of Borel probability measures on $\mathbb{R}^d$, and $v : (0,T) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a Borel velocity vector field (we always use the notation $v_t(x) := v(t,x)$) satisfying the finite $p$–energy condition:

$$E_p(v) := \int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p \, d\mu_t(x) \, dt < +\infty, \quad \text{for some } p > 1, \quad (2.1)$$

to be a solution of (1.6) and the equation has to be intended in the sense of distributions, i.e. $(\mu,v)$ satisfies the continuity equation in the distributions sense.

Note that, the assumption that $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ could be replaced by $\mu_t \in \mathcal{P}_p(\mathbb{R}^d)$, $p > 1$ (the set of Borel probability measures on $\mathbb{R}^d$ with finite $p$–moment, i.e. $\int_{\mathbb{R}^d} |x|^p \, d\mu(x) < +\infty$). A simple approximation argument by regular curves shows that: if $(\mu,v)$ satisfies the continuity equation (1.6) with respect to the $L^p$–integrability condition above, then $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ with $\mu_0 \in \mathcal{P}_p(\mathbb{R}^d)$, $p > 1$ implies $[0,T] \ni t \mapsto \mu_t \in \mathcal{P}_p(\mathbb{R}^d)$ (we refer to Lemma 8.1.9 [17] for the precise notion of solution).

**Remark 2.1** If $\mu_0^s \in \mathcal{P}(\mathbb{R}^d)$, defining the narrowly continuous family of measures $[0,T] \ni t \mapsto \mu_t := X_t(s,\cdot)_# \mu_0^s$, it is immediate to check that $\mu_t \in \mathcal{P}_p(\mathbb{R}^d)$, for every $t \in [0,T]$, solving
the continuity equation (1.6) under the condition $\mu_s = \mu_0^s$ with respect to a Borel vector field $v_t$ satisfying (2.1); in this case we say that $X$ is a flow associated with $(\mu_t, v_t)$.

### 2.1.1 Probabilistic solutions of the continuity equation

As we have already seen in Chapter 1, without assuming any regularity on $v_t$ we do not know anything about the existence and the uniqueness of a flow associated to $(\mu_t, v_t)$. We are mainly interested in the properties of flows out of the smooth setting. We remark that the applicability of the smooth context of the Cauchy–Lipschitz theory, namely in the cases presented in Section 1.2 (for instance Lipschitz regularity, one–sided Lipschitz condition, Osgood condition), is in fact very limited. In general, out of smooth setting of the velocity vector field $v$ satisfying only the condition (2.1), the flow $X(t, \cdot)$ associated to $v$ is not defined, and the representation (1.29) does not make sense.

In the applications, it is important to work with non regular vector fields $v$, satisfying only (2.1). In this case, the flow $X_t$ associated to $v$ is not defined, in general, and the representation (1.29) does not make sense. Nevertheless, another type of representation, strictly linked to the previous one, is possible (see Theorem 8.2.1 in [17]): every continuous time dependent Borel probability solution $t \mapsto \mu_t$ of the continuity equation (1.6) with the vector field $v$ satisfying (2.1) is representable by means of a Borel probability measure $\eta$ on the space $\Gamma_T := C([0, T]; \mathbb{R}^d)$ of continuous maps from $[0, T]$ into $\mathbb{R}^d$ endowed with the sup norm. The measure $\eta$ is concentrated on the set of the curves $t \mapsto X_t$ such that $X_t$ is an integral solution (1.5) of (1.3) and $\dot{X} \in L^p(0, T; \mathbb{R}^d)$.

**Definition 2.2 (Generalized Flow)*** Let $v : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d$ be a Borel vector field. The probability measure $\eta \in \mathcal{P}(\mathbb{R}^d \times \Gamma_T)$ is said to be a generalized flow of $v$ if $\eta$ is concentrated on the trajectories $y \in \Gamma_T$ which are absolutely continuous integral solutions of the DE (1.3) starting from $x$, for $\mu_0$–a.e. $x \in \mathbb{R}^d$. We say that a generalized flow $\eta$ is regular if there exists $C \geq 0$ satisfying $e_{t\#} \eta[\mathbb{R}^d] \ll C\mathcal{L}^d$ for every $t \in [0, T]$, i.e.

$$
\int_{\mathbb{R}^d \times \Gamma_T} \xi(e_t(y)) \, d\eta(x, y) \leq C \int_{\mathbb{R}^d} \xi(z) \, dz \quad \text{for every } \xi \in C^0_0(\mathbb{R}^d), \ t \in [0, T],
$$

(2.2)
where \( e_t : \Gamma_T \rightarrow \mathbb{R}^d \) denotes the evaluation map, defined by \( e_t(y) := y(t) \), and the pushforward is defined by \( e_t#\eta(B) := \eta(\{y \in \Gamma_T : y(t) \in B\}) \) for every Borel set \( B \) of \( \mathbb{R}^d \).

**Remark 2.3** Observe that the definition of the evaluation maps \( e_t \) does not contain \( x \), so due to the first statement the measure \( \eta \) in the previous definition can also be identified with a measure \( \sigma \) in \( \Gamma_T \) whose projection on \( \mathbb{R}^d \) via the map \( e_0 : y \mapsto y(0) \) is \( \mu_0 \) and whose corresponding disintegration \( \sigma = \int_{\mathbb{R}^d} \sigma_x \, d\mu_0(x) \) is made by probability measures \( \sigma_x \) concentrated on solutions of the ODE (1.1) starting from \( x \) at \( t = 0 \). In this case the left hand side of (2.2) takes the simpler equivalent form \( \int_{\Gamma_T} \xi(y(t)) \, d\sigma(y) \), i.e.

\[
\int_{\mathbb{R}^d \times \Gamma_T} \xi(y(t)) \, d\eta(x, y) := \int_{\mathbb{R}^d} \int_{\Gamma_T} \xi(y(t)) \, d\eta_x(y) \, d\mu_0(x) \quad \text{for every } \xi \in C_0^b(\mathbb{R}^d), \, t \in [0, T].
\]

In order to better understand the meaning of the probabilistic representation we recall that, from formula (1.29) and from the characterization of the pushforward in (3.3), it follows that, when there is a unique flow \( X_t(x) \) associated to the vector field \( v \), the only solution of the continuity equation with initial data \( \mu_0 \in \mathcal{P}(\mathbb{R}^d) \) is the measure \( \mu_t \) characterized by

\[
\langle \xi, \mu_t \rangle = \int_{\mathbb{R}^d} \xi(X_t(x)) \, d\mu_0(x) \quad \text{for every nonnegative } \xi \in C_c(\mathbb{R}^d).
\]

(2.3)

For every \( x \in \mathbb{R}^d \) let us consider a probability measure \( \eta_x \in \mathcal{P}(\Gamma_T) \) concentrated on the trajectories \( y \in \Gamma_T \) which are absolutely continuous integral solutions of the ODE (1.21) for \( t \in [0, T] \). All the families \( \{\eta_x\}_{x \in \mathbb{R}^d} \) in the following discussions is weakly measurable, i.e. for every function \( \Xi \in C_b(\Gamma_T) \) and every \( x \in \mathbb{R}^d \) the map \( x \mapsto \langle \Xi, \eta_x \rangle = \int_{\Gamma_T} \Xi(y) \, d\eta_x(y) \) is measurable.

**Definition 2.4 (Probabilistic solution)** The probabilistic solution induced by the generalized flow \( \eta \) is a Borel probability measure \( \mu^\eta_t \in \mathcal{P}(\mathbb{R}^d) \), for \( t \in [0, T] \), defined as follows:

\[
\langle \xi, \mu^\eta_t \rangle := \int_{\mathbb{R}^d} \int_{\Gamma_T} \xi(y(t)) \, d\eta_x(y) \, d\mu_0(x) \quad \text{for every } \xi \in C_0^b(\mathbb{R}^d), \, t \in [0, T].
\]

(2.4)

Notice that, the definition of the evaluation maps \( e_t \) does not contain \( x \), and so also the basic “representation” formula (2.4) defining a solution \( \mu^\eta_t = e_t#\eta \) of the continuity equation
(1.6) for every \( t \in [0, T] \) does not contain \( x \): for a discussion of this matter see Remark 2.3. When this relation between \( \eta \) and \( \mu_t \) holds, we say that \( \eta \) represents the curve \( \mu_t \). We can regard the probabilistic solution of Definition 2.4 as a “superposition principle” of (1.6): considering \( \mu_t \) is a positive locally finite measure–valued solution of the continuity equation, if there is more than one solution to the ODE, then we define our “averaged pushforward” by substituting the quantity \( \xi(X_t(x)) \) with the average \( \int_{\Gamma_T} \xi(y(t)) \, d\mu_t(y) \). It is not difficult, arguing as in the verification of (1.29), to check that (2.4) defines a solution of the continuity equation, for every family \( \{ \eta_x \}_{x \in \mathbb{R}^d} \) as above (see for instance [12], [41]).

In the case when the generalized flow \( \eta \) is the pushforward under \( x \mapsto X^x \) of \( \mu_0 \) (here we are considering \( X \) as a function mapping \( x \in \mathbb{R}^d \) into the solution curve \( t \mapsto X_t(x) \) in \( \Gamma_T \)) we see that the probability measures \( \mu_t^q \) implicitly defined by (2.4) simply reduce to the stranded ones considered in the smooth setting, i.e. \( \mu_t^q = X_t^\# \mu_0 \). Indeed, in the case of a smooth bounded vector field when for every \( x \in \mathbb{R}^d \) the solution of the ODE (1.3) starting from \( x \) is unique, a particular class of generalized flows is the one generated by transporting the initial measure along the integral lines of the flow \( \eta = \int_{\mathbb{R}^d} \delta_{X(. , x)} \, d(\epsilon_0^\# \eta)(x) \), i.e. the only admissible measures \( \eta_x \) in (2.4) are, indeed, the Dirac masses for \( \mu_0 \)-a.e. \( x \in \mathbb{R}^d \).

**Proposition 2.5** Let \( v : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d \) be a bounded vector field and let \( \eta \) be a generalized flow. If the continuity equation (1.6) with any initial datum \( \mu_0 \ll \mathcal{L}^d \) is well–posed and \( \mu_t \) are represented in terms of \( \eta \) in the sense of Definition 2.4, then \( \eta \) is a Dirac mass for \( \mathcal{L}^d \)-a.e. \( x \in \mathbb{R}^d \).

**Sketch of a Proof.** The proof is achieved by a kind of dyadic decomposition of \( \mathbb{R}^d \): if \( \eta \) happen not to be Dirac masses in a set with positive \( \mathcal{L}^d \)-measure, we are able to find two solutions of the continuity equation with the same initial value \( \mu_0 \ll \mathcal{L}^d \), that become orthogonal before time \( T \) (see for example Lemma 15 [12] or Proposition 6.4.3 [41] for a complete proof).

Finally, denoting \( \eta_x = \delta_{X(. , x)} \) and using the notation of probabilistic curve we can give an alternative interpretation of (2.3) as following

\[
\langle \xi, \mu_t^\delta_{X(. , x)} \rangle = \int_{\mathbb{R}^d} \int_{\Gamma_T} \xi(y(t)) \, d\delta_{X(. , x)}(y) \, d\mu_0(x) = \int_{\mathbb{R}^d} \xi(X_t(x)) \, d\mu_0(x),
\]

so in this case we have recovered to the “deterministic” formula (2.3). \( \square \)
2.1.2 Probabilistic representation of the measure–valued solutions

The probabilistic representation says that, for positive solutions, this construction can be reversed: every positive measure–valued solution $\mu_t$ of (1.6) can be realized as a probabilistic solution $\mu_t^\eta$ for some $\eta$ as above. The probabilistic representation for solutions of the homogeneous continuity equation is provided by Theorem 8.2.1 in [17], and for the convenience of the reader we recall the precise statement of this result (that gave us the initial hint for our discussion to the so–called a regular Lagrangian flow). We refer also to a slightly generalization result for a superposition principle in Section 3 of [6], where the author is needed to deal with bounded solutions of the continuity equation so he considered there measures $\mu_t$ which are only a positive locally finite and the vector field satisfies the bounded condition (1.7). See also [12], [13], for an extension to the vector field satisfies the global summability assumption (1.30). In the proof of this case, authors use the narrow (or, weak) convergence of positive measures, i.e. the convergence with respect to the duality with continuous and bounded functions, and the easy implication in Prokhorov compactness theorem: any tight and bounded family $\mathcal{F}$ in $\mathcal{M}_+(\mathcal{X})$ is (sequentially) relatively compact with respect to the narrow convergence. Remember that tightness means: for any $\varepsilon > 0$ there exists $K \subset \mathcal{X}$ compact s.t. $\mu(\mathcal{X} \setminus K) < \varepsilon$ for every $\mu \in \mathcal{F}$. A necessary and sufficient condition for tightness is the existence of a coercive functional $\Phi : \mathcal{X} \to [0, +\infty]$ such that $\int \Phi \, d\mu \leq 1$ for any $\mu \in \mathcal{F}$.

**Theorem 2.6 (Probabilistic representation)** Let $\mu_t : [0,T] \to \mathcal{P}(\mathbb{R}^d)$ be a narrowly continuous Borel measures solving the continuity equation (1.6) for a suitable Borel vector field $v$ satisfying (2.1) for some $p > 1$. Then, $\mu_t$ is a probabilistic solution, i.e. there exists a generalized flow $\eta \in \mathcal{P}([0,T])$ is concentrated on the set of $(x,y)$ with $y \in AC^p(0,T;\mathbb{R}^d)$ is a solutions of the ODE (1.1) for a.e. $t \in (0,T)$, such that $\mu_t = \mu_t^\eta$ for every $t \in [0,T]$. Conversely, every generalized flow $\eta$ satisfying

$$\int_0^T \int_{\mathbb{R}^d \times \Gamma_T} |v_t(y(t))| \, d\eta(x,y) \, dt < +\infty,$$

induces, via (2.4), a solution of the continuity equation (1.6), with $\mu_0 = y(0)\#\eta$. 
Notice that existence of solutions to the continuity equation implies, by the probabilistic representation, existence of the generalized flow and the solutions are probabilistic solutions. For a proof of Theorem 2.6 see for instance [17], or [87] for the probabilistic representation in the case of nonhomogeneous continuity equation. The interested reader may consult Section 4 of [12] or Section 6.2 of [41] for a complete proof in a more general context.

The importance of the probabilistic representation also relies in the fact that it will allow, using truncations and restrictions of the measures \( \eta_x \), several manipulations of solutions of the continuity equation: these constructions are not immediate at the level of the PDE, but they are extremely useful in various occasions, as we will see in the proof of Theorem 2.10.

### 2.2 ODE uniqueness vs PDE uniqueness

The importance of such a probabilistic representation is that, under suitable hypothesis on the vector field \( v \), it gives uniqueness results for positive solutions of the continuity equation once we know, in some way, uniqueness of the characteristic curves associated to the vector field \( v \). This happens for example in the well known case of Lipschitz (w.r.t. the space variable) vector fields, or of vector fields satisfying an Osgood type condition. The proof of next theorem, which presents a very general criterion relating the pointwise uniqueness for the ODE (1.1) with the uniqueness for positive measure–valued solutions to the PDE (1.6), is based on tools developed in [12], [17], [41]. This criterion is, indeed, followed from the probabilistic representation concept. A crucial observation is that whenever suitable measures is considered in the space of continuous maps, allowing for the probabilistic representation of trajectories, which roughly speaking says that these measures are generalized flows such that every probability measure solves the continuity equation (1.6) can be obtained as a probabilistic of solutions obtained via propagation along characteristics.

Notice that, in the case of a finite dimensional space, \((\mu_t, v_t)\) satisfies the continuity equation (1.6) if and only if \( \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \) in \( D'((0, T) \times \mathbb{R}^d) \). Note also in order to give a meaning to the product \( \mu v \) when \( \mu \) is a measure, we assume \( v \) to be defined everywhere in \((0, T) \times \mathbb{R}^d\). Note also
Theorem 2.7 (Pointwise uniqueness) Let $B \subset \mathbb{R}^d$ be a Borel set. Then, the ODE (1.1) admits a unique solution for every initial point $x \in B$ if and only if the PDE (1.6) admits a unique positive measure–valued solution for every initial data $\mu_0$ which is a positive measure concentrated on $B$, i.e. such that $\mu_0(\mathbb{R}^d \setminus B) = 0$.

Proof. First, let’s assume that $\mu_t$ is a positive measure–valued solves the continuity equation with initial data $\mu_0$. Applying the notation of probabilistic representation (see Theorem 2.6 above) we deduce that $\mu_t^\eta = \mu_t$ with the generalized flow $\eta \in \mathcal{P}(\mathbb{R}^d \times \Gamma_T)$ is concentrated on the absolutely continuous integral solutions (1.5) of the ODE (1.1) starting from $x$, for every point $x \in B$. But the assumption of a uniqueness property to solutions of the ODE (1.1) for every initial point $x \in B$, precisely means that, for every $x \in B$, the measures $\eta_x$ are unique. Hence, according to Proposition 2.5, for every $x \in B$ the measure $\eta_x$ are a Dirac masses supported on the unique trajectory starting from $x$, and eventually this gives an explicit formula for the solution $\mu_t$ of the PDE (1.6), which is therefore unique.

The proof of the converse direction is rather easy. Assume that for some $x \in \mathbb{R}^d$ there exist two different solutions $y(t)$ and $\tilde{y}(t)$ of the ODE (1.1) starting from $x$. Then consider $\mu = \delta_{y(t)}$ and $\tilde{\mu} = \delta_{\tilde{y}(t)}$. We clearly have $\mu_0 = \tilde{\mu}_0 = \delta_x$. It is readily checked that $\mu_0$ and $\tilde{\mu}_0$ are solutions of the continuity equation, but since they are different we are violating the reverse assumption. \hfill \Box

The result of Theorem 2.7 is very sharp and elegant, but its applicability is in fact very limited. On one hand, pointwise uniqueness for the ODE is known only under very strong regularity assumptions on the vector field, namely under assumptions of Lipschitz regularity, one–sided Lipschitz condition, Osgood condition. On the other hand, uniqueness for the continuity equation is known only for particular classes of solutions, typically for solutions which are bounded functions. It is reasonable that this kind of “weaker PDE uniqueness” should reflect into a weaker notion of uniqueness for the ODE: this leads to a new concept of solutions, which is presented in the next subsection. In [17] the implication of how to get uniqueness of the PDE (1.6) from uniqueness of the ODE (i.e. to what extent (1.20) has a unique solution for $\mathcal{L}^d$–almost every initial condition $x$) was also studied, but we will not consider this aspect of the problem during the next subsections.
2.3 Ambrosio’s notion of Lagrangian flows

As we have seen earlier in Section 1.2, the existence and uniqueness of the flow are guaranteed by the classical Cauchy–Lipschitz theorem (we refer to Section 1.2). It turns out that additional regularity of the vector field is inherited by the flow, as for instance regularity of $x \mapsto X(t, x)$. In the case when the low regularity of the velocity vector field $v$ is imposed, the notion of pointwise uniqueness of the flow problem (1.3) is not any more the appropriate one, since it is known only under very strong regularity assumptions on the vector field $v$ (we refer to observations at the end of Subsection 2.2). We need to relax the notion of solution, assuming an additional condition which will ensure uniqueness under the various weak differentiability assumptions presented in the previous sections.

2.3.1 The regular Lagrangian flow

The study of (1.3) out of the smooth context is of great importance (for instance, in view of the possible applications to conservation laws or to the theory of the motion of fluids) and has been studied by several authors. What can be said about the well–posedness of (1.3) when $v$ is only in some class of weak differentiability? We remark from the beginning that no generic uniqueness result (i.e. for a.e. initial datum $x$) is presently available. This question can be, in some sense, “relaxed” (and this relaxed problem can be solved, for example, in the Sobolev [56] or $BV$ [6] framework): we look for a canonical selection principle, i.e. a strategy that “selects”, for a.e. initial datum $x$, a solution $X(\cdot, x)$ in such a way this selection is stable with respect to smooth approximations of $v$. This in some sense amounts to redefine the notion of solution: we add some conditions which select a “relevant” solution of our equation. With this new notion of solution we will have uniqueness “in the selection sense”: whenever a smooth sequence $v_h$ is approximated to the vector field $v$, one can obtain that the classical flows $X_h$ associated to them converge to the chosen flow for $v$. This concept of solution is encoded in the following definition: we consider only the flows such that there are no concentrations of the trajectories, loosely speaking, an “almost everywhere flow which (almost) preserves the Lebesgue measure”. 
**Definition 2.8 (Regular Lagrangian flow)** Let \( v \in L^\infty((0, T) \times \mathbb{R}^d, \mathbb{R}^d) \). We say that a map \( X : (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \) is a regular Lagrangian flow for the vector field \( v \) if

(i) For \( \mathcal{L}^d \)-a.e. \( x \in \mathbb{R}^d \) the map \( t \mapsto X(t, x) \) is absolutely continuous integral solutions of (1.21) for \( t \) in \([0, T] \);

(ii) There is some positive real constant \( C \) independent of \( t \) such that the compressibility condition \( X_t \# \mathcal{L}^d \ll C \mathcal{L}^d \) holds for every \( t \in [0, T] \),

where \( X_t \# \mathcal{L}^d \) denotes the pushforward of the \( d \)-dimensional Lebesgue measure \( \mathcal{L}^d \) via the flow map \( X : (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \), defined according to Definition 3.2.

Notice that, the condition (ii) above meaning that the density will be absolutely continuous with respect to Lebesgue measure and bounded by some positive constant: if \( \mu_t := X_t \# \mathcal{L}^d \) and \( \mu = \rho \mathcal{L}^{1+d} \), then \( \|\rho\|_\infty \leq C \) for some constant \( C > 0 \), which is called the compressibility constant of \( X \). Indeed, this condition can be equivalently reformulated as follows: there exists a constant \( C > 0 \) such that for every \( t \in [0, T] \) and every \( \xi \in C_c(\mathbb{R}^d) \) with \( \xi \geq 0 \) there holds \( \int_{\mathbb{R}^d} \xi \circ X_t(x) \, dx \leq C \int_{\mathbb{R}^d} \xi(x) \, dx \). This means that we require \textit{a priori} (i.e., as a sort of “selection condition” for our notion of solution) a quantitative control on how much the flow compresses \( d \)-dimensional sets. For simplicity, in the following presentation, we shall restrict our attention to those regular Lagrangian flow which exactly preserve the Lebesgue measure (in the smooth context, this corresponds to the condition of \( v \) having zero divergence, thanks to Liouville’s Theorem). We can formulate it by saying that “changes of variable along the flow are performed for free”, that is, \( \int_{\mathbb{R}^d} \xi(X_t(x)) \, dx = \int_{\mathbb{R}^d} \xi(x) \, dx \) for every \( \xi \in C_c(\mathbb{R}^d) \).

With this concept of the regular Lagrangian flow we have the well–posedness of (1.3) in the nonsmooth framework.

### 2.3.2 \( \mathcal{L}_v \)-Lagrangian flows

In this context, Ambrosio [6] has defined a more general concept of solution, the one of \( \mathcal{L}_v \)-Lagrangian flows (see Definition 13 of [12], [13]): we consider a convex class \( \mathcal{L}_v \) of positive measure–valued solutions \( \mu_t \in \mathcal{M}_+(\mathbb{R}^d) \) of the continuity equation (1.6) with vector
field \( \mathbf{v} \), satisfying the monotonicity property:

\[
0 \leq \mu'_t \leq \mu_t \in \mathcal{L}_v \quad \Rightarrow \quad \mu'_t \in \mathcal{L}_v,
\]

whenever \( \mu'_t \) solves (1.6), and satisfies the appropriate integrability condition (1.30). Then, we can define the notion of \( \mathcal{L}_v \)-Lagrangian flow starting from given measure \( \mu_0 \in \mathcal{M}_+(\mathbb{R}^d) \), requiring that the map \( X(\cdot, x) \) is an absolutely continuous in \([0, T] \) and satisfies (1.5) for \( \mu_0 \)-a.e. \( x \in \mathbb{R}^d \) (i.e. it is an integral solution of the DE (1.3)), and that the measures \( \mu_t = X(t, \cdot)\#\mu_0 \) induced via pushforward belong to the class \( \mathcal{L}_v \). Heuristically, \( \mathcal{L}_v \)-Lagrangian flows can be thought as suitable selections of the (possibly nonunique) solutions of the ODE, made in such a way to produce densities in \( \mathcal{L}_v \), see Example 2.9 below for an illustration of this concept. Indeed, the notion of a regular Lagrangian flow is a particular case of \( \mathcal{L}_v \)-Lagrangian flow when a solution be considered in the most natural class \( \mathcal{L}_v := L^\infty([0, T] \times \mathbb{R}) \), since typically the well-posedness results at the PDE level can be shown for bounded solution. Therefore, the regular Lagrangian flow is the “good” notion of solution in this low regularity context: roughly speaking, among the many nonunique ODE flows, we pick the flow which does not concentrate trajectories (this flow also enjoys suitable stability properties). We shall enter in more details of this theory in the next section.

**Example 2.9** If we consider again the square root example (see Example 1.3), then solutions of the DE (1.4) are not unique for \( x = -c^2 < 0 \). Indeed, they reach the origin in time \( 2t_\varepsilon \), where they can stay for an arbitrary time \( T \), then continuing as \( X(t, x) = (t - T - 2\varepsilon)^2 \). Moreover, if we consider, for instance, the Lipschitz approximation (that could easily be made smooth) of \( v(y) = 2\sqrt{|y|} \) by

\[
v_\varepsilon(y) = \begin{cases} 
2\sqrt{|y|} & \text{if } y \in (-\infty, -\varepsilon^2] \\
\varepsilon & \text{if } y \in [-\varepsilon^2, \lambda_\varepsilon - \varepsilon^2] \\
2\sqrt{y - \lambda_\varepsilon + 2\varepsilon^2} & \text{if } y \in [\lambda_\varepsilon - \varepsilon^2, +\infty),
\end{cases}
\]

with \( \lambda_\varepsilon - \varepsilon^2 > 0 \), then, solutions of the approximating DE’s starting from \(-c^2\) reach the value \(-\varepsilon^2\) in time \( t_\varepsilon = 2(c - \varepsilon) \) and then they continue with constant speed \( \varepsilon \) until they reach
\( \lambda_\varepsilon - \varepsilon^2 > 0 \) in time \( T_\varepsilon = \lambda_\varepsilon / \varepsilon \). Then, they continue as \( \lambda_\varepsilon - 2\varepsilon^2 + (t - t_\varepsilon - T_\varepsilon)^2 \). Choosing \( \lambda_\varepsilon = \varepsilon T \), with \( T > 0 \), by this approximation we select the solutions that don’t move, when at the origin, exactly for a time \( T \).

Other approximations, as for instance \( v(y) = \sqrt{\varepsilon + |y|} \), select the solutions that move immediately away from the singularity at \( y = 0 \). Among all possibilities, this family of solutions \( X(t, x) \) is singled out by the property that \( X(t, \cdot)\#\mathcal{L}^1 \) is absolutely continuous with respect to \( \mathcal{L}^1 \), so that no concentration of trajectories occurs at the origin. To see this fact, notice that we can integrate in time the identity

\[
0 = X(t, \cdot)\#\mathcal{L}^1(\{0\}) = \mathcal{L}^1(\{x \in \mathbb{R} : X(t, x) = 0\})
\]

and use Fubini’s theorem to obtain

\[
0 = \int_{\mathbb{R}} \mathcal{L}^1(\{t : X(t, x) = 0\}) \, dx.
\]

Hence, for \( \mathcal{L}^1 \)-a.e. \( x \), the trajectory \( X(\cdot, x) \) does not stay at 0 for a strictly positive set of times.

We notice that, there is a similar version if only certain \( W^{1, p} \)-regularity on \( v \) is assumed. Indeed, an alternative approach to the theory of regular Lagrangian flows, based on some quantitative estimates along the flow itself, is provided in [45] and gives well-posedness for vector fields belonging to the Sobolev space \( W^{1, p} \) with \( p > 1 \). For a detailed exposition of all these results we refer to [12], [42]. We also mention, in the Sobolev case, the almost (minor improvements are available [12], [41]) state-of-the-art uniqueness result covers only velocities

\[
v \in L^\infty((0, T) \times \mathbb{R}^d; \mathbb{R}^d) \cap L^1([0, T]; W^{1, p}(\mathbb{R}^d; \mathbb{R}^d)) \quad \text{for some } p > 1. \tag{2.5}
\]

It is proved, in Theorem 2.10 below, that if \( v \) is this regular, then its regular Lagrangian flow, if it exists, then it is unique.

We refer the reader to the scientific literature [10] for further studies on maximal regular flows of nonsmooth vector fields. In that paper Ambrosio, Colombo and Figalli provided a complete analogy between the Cauchy–Lipschitz and the DiPerna–Lions theories for ODEs,
by developing a local version of the DiPerna-Lions theory \[56\]. More precisely, they proved existence and uniqueness of a maximal regular flow for the DiPerna–Lions theory using only local regularity and summability assumptions on the vector field, in analogy with the classical theory, which uses only local regularity assumptions. They also studied the behaviour of the ODE trajectories before the maximal existence time. Unlike the Cauchy–Lipschitz theory, this behaviour crucially depends on the nature of the bounds imposed on the spatial divergence of the vector field. In particular, a global assumption on the divergence is needed to obtain a proper blow–up of the trajectories.

### 2.3.3 Well–posedness of Lagrangian flows

In this subsection we present a sketch of the derivation of the results of existence, uniqueness and stability for the regular Lagrangian flow deduced from the well–posedness in the class of bounded solutions for the continuity equation. With the same techniques also existence and stability results can be proved. This abstract passage is due to Ambrosio; we refer to \[6\] for the original approach in the $BV$ case and to \[7\] and \[12\] for the formalization of the argument in the general case. We notice that these results, in the Sobolev framework, have been already obtained by DiPerna and Lions in \[56\] with different techniques.

This approach is strongly based on the notion of probabilistic solution we presented in Subsection 2.1: starting from the “generalized flow” given by the measures $\eta \in \mathcal{P}(\mathbb{R}^d \times \Gamma_T)$ we perform various constructions at the level of measure–valued solutions of the continuity equation; at that point the PDE well–posedness comes into play, allowing to deduce results about the measures $\eta_x \in \mathcal{P}(\Gamma_T)$ for $x \in \mathbb{R}^d$, roughly speaking obtaining that the generalized flow is in fact a regular Lagrangian flow, since $\eta_x$ selects a single trajectory for $\mathcal{L}^d$–a.e. $x \in \mathbb{R}^d$.

**Theorem 2.10 (Existence and uniqueness of the regular Lagrangian flow)** Let $v \in L^\infty([0, T] \times \mathbb{R}^d; \mathbb{R}^d)$ and assume that the continuity equation \((1.6)\) has the uniqueness property in $L^\infty([0, T] \times \mathbb{R}^d)$. Then, the regular Lagrangian flow associated to $v$, if it exists, is unique. Assume in addition that the continuity equation \((1.6)\) with initial data $\mu_0 = \mathcal{L}^d$ has a positive solution in $L^\infty([0, T] \times \mathbb{R}^d)$. Then, we have existence of a regular Lagrangian flow relative
to \( \mathbf{v} \).

This theorem is a simple consequence of Proposition 2.5, which is an uniqueness result in the wider class of the “multivalued solutions” given by the measures \( \{ \eta_x \}_{x \in \mathbb{R}^d} \). With a little abuse of terminology, in the statement of the theorem we write “\( \mu_t \in L^\infty \)” meaning that we require that \( \mu_t \) is an absolutely continuous (with respect to \( L^d \)) measure, whose density is essentially uniformly bounded. With uniqueness property in \( L^\infty([0,T] \times \mathbb{R}^d) \) we mean that, for any initial data \( \mu_0 \in L^\infty(\mathbb{R}^d) \), if \( \mu^1_t, \mu^2_t \in L^\infty([0,T] \times \mathbb{R}^d) \) are two solutions of (1.6) with \( \mu^1_0 = \mu^2_0 = \mu_0 \), then we must have \( \mu^1_t(x) = \mu^2_t(x) \) for \( L^d+1 \)-a.e. \( (t,x) \in [0,T] \times \mathbb{R}^d \).

Notice that, using a simple regularization argument, the existence of a positive solution assumed in the theorem is satisfied for instance if \( \mathbf{v} \) has bounded divergence, i.e. \( \nabla \cdot \mathbf{v} \ll L^d \) for \( L^1 \)-a.e. \( t \in [0,T] \).

**Remark 2.11 (Uniqueness of \( L_v \)-Lagrangian flows)** Theorem 2.10 could be generalized as follows: first recalling the notion of \( L_v \)-Lagrangian flow discussed in Subsection 2.3.2. Then, a uniqueness of \( L_v \)-Lagrangian flow starting from \( \mu_0 \) is followed immediately whenever the continuity equation has a uniqueness property in \( L_v \). The proof is very similar to the one we are going to present: the reader is referred to Section 4 of [12], [13] for the details and the modifications needed.

In the remaining part of this section we illustrate the proof of Theorem 2.10. We will show that the proof is a consequence of Proposition 2.5.

**Proof of Theorem 2.10.** The proof of Theorem 2.10 easily comes from Proposition 2.5. Assume that there exist two different regular Lagrangian flows \( \mathbf{X}_1(t,x) \) and \( \mathbf{X}_2(t,x) \) relative to the vector field \( \mathbf{v} \). We first consider for \( L^d \)-a.e. \( x \in \mathbb{R}^d \) the measures \( \eta^1_x = \delta_{\mathbf{X}_1(\cdot,x)} \) and \( \eta^2_x = \delta_{\mathbf{X}_2(\cdot,x)} \). We define \( \eta_x = (\eta^1_x + \eta^2_x)/2 \). It is clear by this definition that the family of measures \( \{ \eta_x \}_{x \in \mathbb{R}^d} \subset \mathcal{P}(\Gamma_T) \) is a generalized flow, i.e. for \( L^d \)-a.e. \( x \in \mathbb{R}^d \) the measure \( \eta_x \) is concentrated on absolutely continuous integral solutions of the ODE (1.1) starting from \( x \). Then, we check that the probabilistic solution \( \mu^{\eta}_t \) induced by the family \( \{ \eta_x \}_{x \in \mathbb{R}^d} \) belongs
to $L^\infty([0, T] \times \mathbb{R}^d)$. Indeed, for every $\xi \in C_c(\mathbb{R}^d)$ we can compute
\[
\langle \xi, \mu_t^{\eta_x} \rangle = \int_{\mathbb{R}^d} \int_{\Gamma_t} \xi(y(t)) \, d\eta_x(y) \, dx = \int_{\mathbb{R}^d} \int_{\Gamma_t} \xi(y(t)) \, d\left( \frac{1}{2} (\delta_{X_1(\cdot, x)} + \delta_{X_2(\cdot, x)}) \right) \, dx \\
= \frac{1}{2} \int_{\mathbb{R}^d} \xi(y) \, d\left(\delta_{X_1(\cdot, x)} + \delta_{X_2(\cdot, x)}\right)(y) + \frac{1}{2} \int_{\mathbb{R}^d} \xi(y) \, d\mathcal{L}^d((\cdot, x)\#\mathcal{L}^d)(y).
\]
This means that $\mu_t^{\eta_x} = (\mu_t^{\eta_x^1} + \mu_t^{\eta_x^2})/2 = (\mathcal{X}_1(\cdot, x)\#\mathcal{L}^d + \mathcal{X}_2(\cdot, x)\#\mathcal{L}^d)/2$, and by condition (ii) in Definition 2.8 we obtain that $\mu_t^{\eta_x}$ is a bounded function, i.e. it belongs to $L^\infty((0, T) \times \mathbb{R}^d)$. Moreover, the measure $\eta_x$ is not a Dirac mass for every $x$ belonging to a set of positive Lebesgue measure, precisely for those $x \in \mathbb{R}^d$ such that $\mathcal{X}_1(\cdot, x) \neq \mathcal{X}_2(\cdot, x)$. This happens in a set of points $x$ of probability measure, since we are assuming that the two regular Lagrangian flows $\mathcal{X}_1$ and $\mathcal{X}_2$ are different. Thus the family $\{\eta_x\}_{x \in \mathbb{R}^d} \subset \mathcal{P}(\Gamma_T)$ defined above contradicts the result of Proposition 2.5, and the uniqueness result of Theorem 2.10 follows.

We now pass to the proof of the second statement Theorem 2.10. To prove the existence of the regular Lagrangian flow we assume that $\mu_t \in L^\infty((0, T) \times \mathbb{R}^d)$ is a probability solution of the continuity equation (1.6) with initial data $\mu_0 = \mathcal{L}^d$. Noticing that the assumptions of Proposition 2.6 are satisfied, we can apply the probabilistic representation to deduce that $\mu_t = \mu_t^{\eta_x}$ for some family $\{\eta_x\}_{x \in \mathbb{R}^d} \subset \mathcal{P}(\Gamma_T)$, with $\eta_x$ concentrated on absolutely continuous integral solutions of the ODE (1.1) starting from $x$, for $\mathcal{L}^d$-a.e. $x \in \mathbb{R}^d$. Since the continuity equation (1.6) has a unique solution $\mu_t \in L^\infty((0, T) \times \mathbb{R}^d)$ we apply Proposition 2.5 to deduce that $\eta_x$ is a Dirac mass for $\mathcal{L}^d$-a.e. $x \in \mathbb{R}^d$. Denote by $\mathcal{X}(\cdot, x)$ the element of $\Gamma_T$ on which $\eta_x$ is concentrated, for $\mathcal{L}^d$-a.e. $x \in \mathbb{R}^d$; we check that the map $\mathcal{X}(t, x)$ defined in this way is a regular Lagrangian flow associated to $\mathbf{v}$. Condition (i) in Definition 2.8 is clearly satisfied because $\eta_x$ is concentrated on solutions of the ODE. To check condition (ii) it is enough to notice that $\mathcal{X}(t, \cdot)\#\mathcal{L}^d = \mu_t \in L^\infty([0, T] \times \mathbb{R}^d)$.

As the proof clearly shows, we don’t really need uniqueness of forward characteristics in a pointwise sense, but rather that any probability measure $\eta$ in $\Gamma_T = C^0([0, T]; \mathbb{R}^d)$ concentrated on absolutely continuous solutions of the ODE (1.1) and satisfying $(e_t)\#\eta = \mu_t$ is representable as $\eta = (\mathcal{X}(\cdot, x))\#\mu_0$ (see Theorem 2.6) for some “natural” flow $\mathcal{X}$. More precisely, this assumption could be replaced by the following less technical, but also much
less specific one: if $v_t$ is the unique minimal velocity field associated to $\mu_t$ (see; for example, Theorem 8.3.1 of in [17]) and the ODE (1.1) associated to $v$ admits at most one valued solution for $\mu_0$-a.e. $x \in \mathbb{R}^d$, then there exists a unique forward flow $X : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ associated to $(\mu_t, v_t)$.

In a similar fashion it is possible to show a result of abstract stability: if the continuity equation is well–posed for each approximating vector field $v_h$ then the associated regular Lagrangian flows converge strongly to the regular Lagrangian flow associated to the limit vector field $v$.

**Theorem 2.12 (Stability of the regular Lagrangian flow)** Assume that a sequence $(v_h)$ of vector fields converges in $L^1_{\text{loc}}([0, T] \times \mathbb{R}^d)$ to $v \in L^\infty([0, T] \times \mathbb{R})$ with $\nabla \cdot v \in L^\infty([0, T] \times \mathbb{R}^d)$, that

$$||v_h||_{\infty} + ||\nabla \cdot v_h||_{\infty} \leq C < +\infty. \quad (2.6)$$

and that the continuity equation (1.6) in $L^\infty([0, T] \times \mathbb{R}^d)$ with the velocity field $v$ is well–posed. Then, the regular Lagrangian flows $X_h$ relative to $v_h$ converge strongly in $L^\infty([0, T]; L^1_{\text{loc}}(\mathbb{R}^d))$ to the regular Lagrangian flow $X$ associated to $v$, i.e.

$$\lim_{h \to +\infty} \int_{B_R} \sup_{t \in [0, T]} |X_h(t, x) - X(t, x)| \, dx = 0 \quad \text{for every } R, T > 0.$$

We only give a sketch of the proof, which is closely related to the one of the uniqueness result of Theorem 2.10. We refer the reader to Section 3 of [7] for more details. The proof is achieved by looking at the convergence of the Young measures induced by $X_h$, whose limit induces a generalized flow.

**Sketch of a Proof.** As in the proof of the probabilistic representation we consider for every $h$ the measure $\eta^h \in \mathcal{P}(\mathbb{R}^d \times \Gamma_T)$ associated to the regular Lagrangian flow $X_h$. Using the bounds (2.6) it is possible to show that $\eta^h$ is locally tight. Consider a limit point $\eta \in \mathcal{P}(\mathbb{R}^d \times \Gamma_T)$. Arguing as in a proof of Theorem 8.2.1 [17] (see also Step 6 of a proof of Theorem 6.2.2 [41]) and using the convergence of $v_h$ to $v$ we obtain that $\eta$ is concentrated on trajectories of the ODE (1.21) with vector field $v$. Moreover from the narrow convergence and from (2.6) we get that $\eta$ induces bounded probabilistic solutions to the continuity equation.
(1.6). We can then apply Proposition 2.5 to get that \( \eta \) is in fact concentrated on a graph of a regular Lagrangian flow associated to \( v \). But under our assumptions we already know from Theorem 2.10 that such a regular Lagrangian flow is unique, and this means that we have \( (x, X_h(\cdot, x)) \# \mathcal{L}^d \rightarrow (x, X(\cdot, x)) \# \mathcal{L}^d \) locally narrowly. From this it is possible to deduce the strong convergence of \( X_h \) to \( X \) in \( L^\infty([0, T]; L^1_{loc}(\mathbb{R}^d)) \).

\[ \square \]

### 2.4 DiPerna–Lions’ notion of Lagrangian flows

The approach to the ordinary differential equation due to DiPerna and Lions [56] is slightly different from the Ambrosio [6] axiomatization we already introduced in Definition 2.8. It is based on an observation that we already made in the smooth setting (see Subsection 1.3.3): if the flow \( X_t(s, x) \) is the characteristic at time \( t \), starting from the point \( x \) at time \( s \) (so that \( X_t(x) = X_t(0, x) \)), then \( X_t(\cdot, \cdot) \) formally solve the system of transport equations

\[
\partial_s X_t(s, x) + v_s(x) \cdot \nabla X_t(s, x) = 0
\]

with a vector field \( v \) in \((0, T) \times \mathbb{R}^d \) (see equation (1.24)).

In this way results relative to the PDE can be transferred to the ODE by using notion of the flow \( X(t, x) : (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \) deduced by DiPerna and Lions [56] and that can be summarized by the following three properties (see Remark 6.7 [6]):

(a) The characteristic equation (1.20) is satisfied in the sense of distributions in \((0, T) \times \mathbb{R}^d\);

(b) The pushforward of the Lebesgue measure satisfies \( \mathcal{L}^d(C) \leq X_t(s, \cdot) \# \mathcal{L}^d \leq C \mathcal{L}^d \) for some constant \( C > 0 \);

(c) The semigroup property holds: for \( \mathcal{L}^d \)-a.e. \( x \in \mathbb{R}^d \) we have \( X_t(s, X_s(s', x)) = X_t(s', x) \) for every \( s, s', t \in (0, T) \).

The semigroup property (c) is used in a essential way in [56] to characterize the flow \( X \) and to prove its stability properties. The approach developed in [6] is based on a careful analysis of the measures transported by the flow, and ultimately on the homogeneous continuity equation only. Ambrosio [6] (see also [12], [41]) explained the difference between the two approaches by recalling the following two well–known facts: Up to a redefinition of the flow \( X \) on a negligible set, condition (a) is equivalent to condition (i) in the definition of regular
Lagrangian flow. However, only upper bounds on the measure produced by the flow are imposed in the approach of [6] in Definition 2.8(ii), while some upper bound as in (b) is indeed necessary for an invariant theory. Moreover, the semigroup property (c) above holds implicitly in the approach of [6], as a consequence of the uniqueness of the flow and the other two assumptions, using the following argument: it is in fact not an axiom, but a theorem (see Remark 6.7 of [6]).

In setting of [6] condition (c) can be recovered as a consequence with the following argument: assume to fix the ideas that \( s' \leq s \leq T \) and define \( \tilde{X}_t(x) = X_t(s', x) \) if \( s' \leq t \leq s \) and \( \tilde{X}_t(x) = X_t(s, X_s(s', x)) \) if \( s \leq t \leq T \). It is immediate to check that \( \tilde{X}_t(x) \) is a solution of the ODE in \([s', T] \) for \( \mathcal{L}^d \)–a.e. \( x \in \mathbb{R}^d \) and that \( \tilde{X}_t \# \mathcal{L}^d \) is bounded by \( C^2 \mathcal{L}^d \), i.e. the compressibility condition \( \tilde{X}_t \# \mathcal{L}^d \leq L^2 \mathcal{L}^d \) holds for \( \tilde{X}_t \); where \( L \) is a compressibility constant for the regular Lagrangian flow \( X \). Moreover by the uniqueness result in Theorem 2.10 (with \( s' \) as initial time) we obtain that \( \tilde{X}_t(x) = X_t(s', x) \), and this means that the semigroup property (c) holds.

In this framework we also mention the recent result by Hauray, Le Bris and Lions [76] in which uniqueness of this notion of flow (for Sobolev vector fields) is shown directly at the ODE level, with a very simple argument.
Chapter 3

Theory of optimal transportations

The aim of this chapter is to recall some well–known facts that shall be needed later on. The presentation has therefore been tailored with a further use in mind, and proofs are only given when they are either very short or of a special interest. Notations are also set here. For a general introduction to optimal transportation, the reader should rather refer to Villani’s books [115], [116] or Santambrogio’s book [104]. For a more abstract and more general exposition, see also the monograph by Ambrosio, Gigli, and Savaré [17].

The optimal transport problem was first formulated by the French mathematician Gaspar Monge in 1781 [89] when the came up with the question of optimizing the transport of a sandpile to an excavation of equal volume by minimizing the total cost. Then, it has relaxed by Leonid V. Kantorovich in 1942 [78], [79]. Indeed, after almost two centuries Kantorovich solved a weaker version of the original Mong problem by relaxing it to a more manageable, linear problem, which is nowadays still an interesting and challenging problem with applications in different fields of mathematic, e.g. calculus of variations, partial differential equation, kinetic theory, differential geometry of metric measures, infinite dimensional linear programming, functional analysis, mathematical economics and in probability theory and statistics. The difference between Monge’s and Kantorovich’s approaches is that Monge’s formulation does not allow the splitting of mass, mainly due to its “relentless” nonlinearity [58], [115], [116] whereas Kantorovich’s approach does.

In this chapter, we present the optimal transport theory, or the so-called the Monge–Kantorovich problem. In Section 3.1, we give a brief introduction to the Monge problem.
In particular, we introduce the optimal transport formulations, which started two centuries ago with Monge’s work on “des remblais et déblais” [89]. This engineering problem consists in minimizing the transport cost between two given mass densities. Section 3.2 is devoted to the relaxation that Kantorovich [78] did of the original Monge problem, and its dual problem, which was in the 1940’s when he interpreted it as an economic equilibrium. In subsection 3.2.2, we present a statement of the primal optimization problem. Subsection 3.2.2 is devoted to the existence of optimal transport plans and couplings. In subsection 3.2.3, we introduce a dual optimization problem. In Section 3.3, we present a regularity of Brenier’s map and its applications. In Subsection 3.3.1, we present existence and uniqueness of optimal transport maps, due to Brenier [34]: In 1991, he succeeded in proving the equivalence between the Monge’s and Kantorovich’s problems in the quadratic cost and under some restrictions on the objects involved. Subsection 3.3.2 is devoted to recall briefly the classical higher regularity theory for solutions to the Monge–Ampère equation. We will not give proofs of the results stated there but just sketch the main ideas behind them. In Subsection 3.3.3, we present the notation of polar factorization and rearrangements of optimal maps, which will be the central tool used in the proof of main results on this work. In the last few lines will be concerned on important results of the so-called measure–preserving maps satisfying each of the flow equation (1.3) and the transport equation (1.17), the most remarkable case being exactly those of the monotone rearrangement of transport map on the real line, the so–called Brenier’s “optimal” map.

3.1 Transport problem in the classical framework

The classical mass transportation problem goes back more than two centuries to a geometer, Gaspard Monge [89]: in 1781 he wanted to move at minimal expense, in 3–dimensional Euclidean space, a heap of a rubble “déblais” to build up a mound or fortification “remblais” minimizing the cost. In the modern optimal mass transportation problem, Monge optimization problem is formulated within Borel measures $\mu_i$, $i = 1, 2$ on Polish spaces $X_i$ (that is, complete, separable metric spaces), respectively, equipped with Borel $\sigma$–algebra.

Mathematically, Monge proposed in [89] a model to describe the work necessary to move
a mass distribution $\mu_1 = \rho_1 \, dx_1$ into a final destination $\mu_2 = \rho_2 \, dx_2$ defined on the measurable (Borel) spaces $\mathcal{X}_1, \mathcal{X}_2$, respectively, such that the densities $\rho_i \geq 0$, $i = 1, 2$, with $\int_{\mathcal{X}_i} \rho_i(x_i) \, dx_i = 1$ (that is originally Borel probability measures representing the height of soil and the depth of an excavation in Monge’s terminology), once the unitary transportation cost function $c(x_1, x_2)$ (called ground cost), which measures the work to move a unit mass from $x_1 \in \mathcal{X}_1$ into $x_2 \in \mathcal{X}_2$, is given; that is, a nonnegative Borel measurable function $c : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}^+ \cup \{+\infty\}$, over all mass preserving (bijections) Borel maps from $\mathcal{X}_1$ into $\mathcal{X}_2$.

Monge’s objective was to find a Borel map $T : \mathcal{X}_1 \to \mathcal{X}_2$ which transports the mass represented by the measure $\mu_1$ into the mass represented by the measure $\mu_2$, we write $T \# \mu_1 = \mu_2$ if the following holds:

$$\int_{T^{-1}(B)} \mu_1(x_1) \, dx_1 = \int_B \mu_2(x_2) \, dx_2 \quad \text{for any Borel set } B \subset \mathcal{X}_2,$$

a map with and minimizing the quantity

$$M_c(S) := \int_{\mathcal{X}_1} c(x_1, S(x_1)) \, d\mu_1(x_1)$$

overall admissible transportation maps $S : \mathcal{X}_1 \to \mathcal{X}_2$ that push $\mu_1$ forward into $\mu_2$. This means the map $T$ minimizes the total cost of transporting $\mu_1$ into $\mu_2$; that is,

$$\min_{S \# \mu_1 = \mu_2} M_c(S) = \int_{\mathcal{X}_1} c(x_1, T(x_1)) \, d\mu_1(x_1), \quad (3.1)$$

where $c : \mathcal{X}_1 \times \mathcal{X}_2 \to [0, +\infty]$ is some given cost function and the minimum is taken over all Borel maps $S : \mathcal{X}_1 \to \mathcal{X}_2$ with the pushforward operator $S \# \mu_1(B) = \mu_1(S^{-1}(B))$ for every measurable Borel sets $B \subset \mathcal{X}_2$ that is characterized by Definition 3.2 below. We say that $T$ is a transport map since the former “transport” condition, $T \# \mu_1 = \mu_2$, which means that $T$ should transport $\mu_1$ onto $\mu_2$, is satisfied, and when $T$ minimizes the total cost (3.1) we call it an optimal transport map.

Denoting by $\mathcal{T}(\mu_1, \mu_2)$ the set of all transport maps $T : \mathcal{X}_1 \to \mathcal{X}_2$ such that $T$ is measurable and $T \# \mu_1 = \mu_2$. With this notation, the optimal mass transportation problem (3.1)
that was first introduced by G. Monge [89] can be stated in modern terms as follows:

**Problem 3.1 (Monge’s problem)** Given two probability measures \( \mu_i \in \mathcal{P}(\mathcal{X}_i) \), \( i = 1, 2 \) and a cost function \( c : \mathcal{X}_1 \times \mathcal{X}_2 \to [0, +\infty] \), solve

\[
C(\mu_1, \mu_2) := \min_{T \in T(\mu_1, \mu_2)} \left\{ M_c(T) := \int_{\mathcal{X}_1} c(x_1, T(x_1)) \, d\mu_1 \right\}.
\]

(MP)

In other words, find \( T : \mathcal{X}_1 \to \mathcal{X}_2 \) such that \( \mu_2 = T_\# \mu_1 \) and \( \int_{\mathcal{X}_1} c(x_1, T(x_1)) \, d\mu_1 \) is minimal.

As mentioned above, G. Monge [89] formulated his original problem in the 1780s with the simple Euclidean distance in \( \mathbb{R}^3 \) as a cost function. But for such a “ground” cost, the question is particularly difficult. To give an idea: Monge detected that the direction of optimal transport should be along straight lines that would be orthogonal to a family of surfaces. Also he found the no–crossing rule of optimal transport rays. Although this proved to be a key observation, the problem remained mostly unsolved for over a century: his characterization of the transport rays was rigorously proved only a century later, by Paul Appell [22], [22]; and in the 1970s, Vladimir Sudakov [110] claimed to have proved the existence of an optimal mapping, but a point in his demonstration was unconvincing (it was corrected by Luigi Ambrosio in 2000 [4], just after another method had been successfully used by Lawrence C. Evans and Wilfrid Gangbo, with stronger assumptions [60]). For a strictly convex cost, however, things are somewhat easier. At the end of the 1980s, Yann Brenier [35], [34] gave a general answer when the cost function is the squared Euclidean distance, and showed the key role convex functions play in that case. Since, his theorem has been extended to arbitrary, strictly convex cost functions, and for measures defined on a variety of domains; those cases will be studied in the next chapter.

**Definition 3.2 (Pushforward of measures)** Let \( S : \mathcal{X}_1 \to \mathcal{X}_2 \) be a Borel map and \( \mu_1 \) be a Borel probability measure on \( \mathcal{X}_1 \), the pushforward of \( \mu_1 \) (or image measure) through \( S \), denoted by \( S_\# \mu_1 \), is the Borel probability measure \( \mu_2 \) and defined by \( \mu_1(\{x_1 \in \mathcal{X}_1 : (x_1, S(x_1)) \in B\}) \); that is,

\[
\mu_2(\mathcal{B}) = \mu_1 \circ S^{-1}(\mathcal{B}) \quad \text{for all Borel sets } \mathcal{B} \subset \mathcal{X}_2.
\] (3.2)
It is easy to check that this definition is equivalent to the so-called change-of-variables formula
\[ \int_{X_2} \zeta(x_2) d(S#\mu_1)(x_2) = \int_{X_1} \zeta \circ S(x_1) d\mu_1(x_1) \]  
(3.3)
for every bounded (or $S#\mu_1$–integrable) Borel function $\zeta : X_2 \to \mathbb{R}$. Indeed, for all Borel set $B \subset X_2$ and $\chi_B \in L^1(\mu_2)$, we have
\[ \mu_2(B) = \int_{X_2} \chi_B d\mu_2 = \int_{X_1} \chi_B \circ S d\mu_1 = \int_{X_1} \chi_{S^{-1}(B)} d\mu_1 = \mu_1(S^{-1}(B)). \]

### 3.2 Kantorovich relaxation of Monge problem

In the context of optimal mass transportation, a relaxed formulation of Monge’s problem was introduced by Leonid Kantorovitch in 1942 [78], [79]. It is easy to see that in this generality Monge’s formulation (MP) can be ill–posed and the problem of the existence of an optimal transport map is very delicate question and far from being trivial even in the one–dimensional Euclidean space $\mathbb{R}$. Indeed, depending on the probability measures, there might be no transport map sending $\mu_1$ onto $\mu_2$, for instance the case $\mu_1$ is discrete (Dirac mass) but $\mu_2$ is uniform: in this case, there are no transport maps at all and the set $T(\mu_1, \mu_2)$ is empty. The next easy counterexample shows that, when $\mu_i$, $i = 1, 2$ are singular measures, an optimal transport map $T$ that attains the minimum (MP) may not exist.

**Example 3.3** On the real line we consider the probabilities $\mu_1 = \delta_0$ and $\mu_2 = \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_1$. For every map $T : \mathbb{R} \to \mathbb{R}$ we have $T#\mu_1 = T#\delta_0 = \delta_{T(0)}$ $\neq$ $\mu_2$. Then no transport map between $\mu_1$ and $\mu_2$ exists, and so the Monge problem (MP) is meaningless for the “trivial” cost $c(x_1, x_2) = |x_1 - x_2|$. The same situation occurs in general when $\mu_1$ has $\sigma$ atoms and $\mu_2$ has $k$ atoms with $\sigma < k$. When the marginal measures are discrete (that is, finite sums of Dirac masses) and with the same number of atoms, i.e. $\mu_i = \frac{1}{k} \sum_{j=1}^{k} \delta_{x^j_i}$ for $i = 1, 2$, the Monge problem becomes a problem in combinatorial optimization. In fact, a transport map $T$ corresponds to a permutation of the induces and, taking, for instance, $c(x_1, x_2) = |x_1 - x_2|$, the Monge problem becomes
\[ \min \left\{ \sum_{j=1}^{k} |x^j_1 - x^j_2| : \sigma \text{ permutation of } 1, \ldots, k \right\}. \]

Note that, the image measure $T#\mu_1$ always includes an atom of mass at least $\mu_1(\{a\})$.
for every atom \( a \in \mathcal{X}_1 \) of \( \mu_1 \). This implies in particular that measures with atoms cannot be sent through a transport map onto measures without atoms. For these reasons, the absence of atoms is a typical assumption on the starting measure \( \mu_1 \) when one wants to solve the problem of Monge (that is, finding an optimal transport map). This means that one needs some restrictions on the measures \( \mu_1 \) and \( \mu_2 \). We can improve on this by adopting Kantorovich’s formulation of the optimal transportation problem. The major advance on this problem is due to Kantorovitch, who proposed in [78], [79] a notion of relaxed formulation of the optimal transport problem.

### 3.2.1 Statement of the primal optimization problem

To avoid the difficulties related to the existence of optimal transportation maps, in 1942 Leonid Kantorovich proposed [78] a “weak” formulation of the Monge’s problem 3.1 which has the great advantage of being a linear problem. He was interested to find a transport map \( T : \mathcal{X}_1 \to \mathcal{X}_2 \) from \( \mu_1 \) to \( \mu_2 \) which minimizes the transportation cost \( C(\mu_1, \mu_2) \) satisfying (3.1) above. The key is to observe that any map \( S \) pushing \( \mu_1 \) forward into \( \mu_2 \) defines in a unique way a Borel measure \( \gamma_S \) on the product space \( \mathcal{X}_1 \times \mathcal{X}_2 \) such that

\[
\gamma_S(\mathcal{X}_1 \times \mathcal{X}_2) = \mu_1(S^{-1}(\mathcal{X}_1)) = \mu_1(\{x_1 \in \mathcal{X}_1 : S(x_1) \in \mathcal{X}_2\}).
\]

This is, in terms of push forwards, equivalent to

\[
(Id_{\mathcal{X}_1} \times S)_\# \mu_1 = \gamma_S;
\]  

(3.4)

where the map \( Id_{\mathcal{X}_1} \times S : \mathcal{X}_1 \to \mathcal{X}_1 \times \mathcal{X}_2 \) is defined by \( Id_{\mathcal{X}_1} \times S(x_1) = (x_1, S(x_1)) \). Note that

\[
\gamma_S(B_1 \times \mathcal{X}_2) = \mu_1(B_1), \quad \gamma_S(\mathcal{X}_1 \times B_2) = \mu_2(B_2) \quad \text{for every Borel sets } B_i \subset \mathcal{X}_i, \ i = 1, 2. \quad (3.5)
\]

Consequently, \( \gamma_S \) is a finite Borel probability measure on the product space \( \mathcal{X}_1 \times \mathcal{X}_2 \) having first marginal equal to \( \mu_1 \) and second marginal equal to \( \mu_2 \) (and, implicitly, same total mass
as $\mu_1$ and $\mu_2$). Since

$$M_\zeta(S) = \int_{\mathcal{X}_1} \zeta(x_1, S(x_1)) \, d\mu_1(x_1) = \int_{\mathcal{X}_1 \times \mathcal{X}_2} \zeta(x_1, x_2) \, d\gamma_S(x_1, x_2) =: K_\zeta(\gamma_S) \quad (3.6)$$

for every continuous function $\zeta \in C(\mathcal{X}_1 \times \mathcal{X}_2)$, we may take $\zeta(x_1, x_2) = c(x_1, x_2)$ (the cost function on $\mathcal{X}_1 \times \mathcal{X}_2$) to infer that Monge’s problem is equivalent to finding optimal map $T$ such that

$$K_c(\gamma_T) = \min_{\gamma_S} K_c(\gamma_S) \quad (3.7)$$

among all Borel probability measures $\gamma_S$ constructed as before. Any minimizer of problem (3.7) will be called optimal transport plan for $c$ and possibly denoted also as $C(\mu_1, \mu_2)$. Therefore, this time we widen the class $\Gamma(\mu_1, \mu_2)$ of transfer plans $\gamma$ by including all the Borel probability measures on the product space $\mathcal{X}_1 \times \mathcal{X}_2$ with marginals $\mu_i$, $i = 1, 2$, not just the ones concentrated on the graphs of transfer maps $S$; that is, for any couple of measures $(\mu_1, \mu_2) \in \mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2)$, we define

$$\Gamma(\mu_1, \mu_2) = \left\{ \gamma \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2) : \mu_i = \pi^i_# \gamma, \ i = 1, 2 \right\}, \quad (3.8)$$

where $\pi^i : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}_i$, $i = 1, 2$ are the canonical projections maps on the components and $\mathcal{P}(\mathcal{X}_i)$ denotes to the family of the Borel probability measures $\mu_i$ on $\mathcal{X}_i$; respectively, having finite first moments. These probability measures over $\mathcal{X}_1 \times \mathcal{X}_2$ are an alternative way to describe the displacement of the particles of $\mu_1$: instead of saying, for each $x_1$, which is the destination $S(x_1)$ of the particle originally located at $x_1$, we say for each pair $(x_1, x_2)$ how many particles go from $x_1$ to $x_2$. It is clear that this description allows for more general movements, since from a single point $x_1$ particles can a priori move to different destinations $x_2$. If multiple destinations really occur, then this movement cannot be described through a map $S$. Notice that the constraints on $\pi^i_# \gamma$, $i = 1, 2$ exactly mean that we restrict our attention to the movements that really take particles distributed according to the distribution $\mu_1$ and move them onto the distribution $\mu_2$. While a general definition of push forward measures $\pi^i_# \gamma$, $i = 1, 2$ are given in Definition 3.2 above, at this point note that the projection conditions on the transport plan $\gamma$; that is, $\pi^i_# \gamma = \mu_i$, $i = 1, 2$, simply mean the statement...
Remark 3.4 Any transport map \( S \) from \( \mu_1 \) to \( \mu_2 \) naturally induces a transport plan \( \gamma = (\text{Id} \times S)_{\#} \mu_1 \). Therefore the set of transport plans always include all transport maps; that is, \( T \) pushes \( \mu_1 \) onto \( \mu_2 \) and the functional takes the form problem (MP), thus generalizing Monge’s problem. Moreover, the measure \( \mu_1 \otimes \mu_2 \in \Gamma(\mu_1, \mu_2) \) so \( \Gamma(\mu_1, \mu_2) \neq \emptyset \). Thus, the set of \( \gamma \) where the minimum is achieved is nonempty and is denoted by \( \Gamma_o(\mu_1, \mu_2) \) in [17].

This generalized problem by L. Kantorovich [78] is much easier to handle than the original one proposed by Monge; for instance in the Monge case we would need existence of at least a map \( T \) satisfying the constraints, but this is not verified (see Example 3.3). On the contrary, there always exist transport plan in \( \Gamma(\mu_1, \mu_2) \) (for instance \( \gamma = \mu_1 \otimes \mu_2 \)). That is why, in the language of Calculus of Variations, Kantorovich’s version is considered a relaxed Monge problem. The Kantorovich relaxation consists in the attempt to solve a new minimization problem which is known as the primal optimization problem and reads

**Problem 3.5 (Kantorovich’s problem)** Given \( \mu_i \in \mathcal{P}(\mathcal{X}_i), i = 1, 2 \) and \( c : \mathcal{X}_1 \times \mathcal{X}_2 \to [0, +\infty] \), we consider the formulation

\[
\mathcal{C}(\mu_1, \mu_2) := \min_{\gamma \in \Gamma(\mu_1, \mu_2)} \left\{ K_c(\gamma) := \int_{\mathcal{X}_1 \times \mathcal{X}_2} c(x_1, x_2) \gamma(dx_1, dx_2) \right\}, \tag{KP}
\]

where \( \Gamma(\mu_1, \mu_2) \) denotes the set of transport plans; that is, a probability measure \( \gamma \) on the product space \( \mathcal{X}_1 \times \mathcal{X}_2 \) whose marginals \( \mu_i, i = 1, 2 \) are coincide with the projections \( \pi^i_{\#} \gamma \), respectively.

This problem really extends Monge’s problem 3.1, for any transport map \( T \) sending \( \mu_1 \) onto \( \mu_2 \) yields a measure \( \gamma \in \Gamma(\mu_1, \mu_2) \), which is \( \gamma = (\text{Id} \times T)_{\#} \mu_1 \), i.e. the only measure on \( \mathcal{X}_1 \times \mathcal{X}_2 \) such that (3.6) holds for every \( \xi \in \mathcal{C}_b(\mathcal{X}_1 \times \mathcal{X}_2) \) and the associated costs of transportation are the same. However, unlike in Monge’s problem, for which there might be no admissible transport map (not to mention an optimal one), in Kantorovich’s version there is always a transport plan, for instance the tensor product \( \mu_1 \otimes \mu_2 \). Even better, it is not difficult to show there is always a solution; see for example Theorem 3.11 below. Thus, the
problem of showing existence of optimal transport maps reduces to proving that an optimal transport plan is concentrated on a graph. It is however clear, from what we already said, that no such result can be expected without additional assumptions on the measures and on the cost.

**Remark 3.6** The generalization above is the standard notion often used to define the optimal transport of a Monge–Kantorovich transportation problem in which Kantorovich suggested to find transfer plan \( \gamma \) instead of looking for a mapping \( T \) such that it minimizes the relaxed transportation cost \((KP)\) over all admissible transport plans from \( \mu_1 \) to \( \mu_2 \). Moreover, the connection between the formulation of Kantorovich \((KP)\) and that of Monge \((MP)\) can be seen by noticing that any transport map \( S \) induces the plan defined by \( (3.4) \) which is concentrated on the graph of \( S \); that is, any optimal transport map \( T \) can be seen as a optimal transport plan \( \gamma \), given by \( \gamma = (\text{Id} \times T) \# \mu_1 \).

Equivalently, in terms of random variables on a nonatomic probability space \((\Omega, \mathcal{P})\), any plan \( \gamma \in \Gamma(\mu_1, \mu_2) \) transference the mass distribution \( \mu_1 \) to \( \mu_2 \) describes the joint distribution of a coupling \((X_1, X_2)\) of \((\mu_1, \mu_2)\). By “coupling” we mean the couple (or pair) \((X_1, X_2)\) of the two random variables \( X_i, i = 1, 2 \) defined on some probability space \((\Omega, \mathcal{P})\) with a joint law marginal distributions \( \text{law}(X_i) = \mu_i, i = 1, 2 \) (this is the same as the measures \( X_i \# \mathcal{P} = \mu_i \)) respectively. One may choose \( \Omega = X_1 \times X_2 \) if \( \mu_i, i = 1, 2 \) are the only laws in the problem.

The condition statement \((3.5)\) can be immediately deduced from the coupling concept. More precisely, if \((x_1, x_2)\) is random and \( \gamma(dx_1, dx_2) \) is a probability measure in two variables \( x_i \in X_i, i = 1, 2 \), its marginal (or projection) on \( X_1 \) is the measure \( X_1 \# \gamma \); respectively, where \( X_1(x_1, x_2) = x_1, X_2(x_1, x_2) = x_2 \); that is, \( \text{law}(x_1, x_2) = \gamma \), then the conditional law of \( x_1 \) given \( x_2 \) is denoted by \( \gamma(dx_1|x_2) \); this is a measurable function \( X_2 \rightarrow \mathcal{P}(X_1) \), obtained by disintegrating \( \gamma \) along its \( x_2 \)-marginal. The conditional law of \( x_2 \) given \( x_1 \) will be denoted by \( \gamma(dx_2|x_1) \). Using conditional distributions we obtain for \( \gamma \in \Gamma(\mu_1, \mu_2) \),

\[
K_c(\gamma) = \int_{X_1} \left( \int_{X_2} c(x_1, x_2) \gamma(dx_2|x_1) \right) \mu_1(dx_1). \tag{3.9}
\]

Any mass at point \( x_1 \) is transported to \( x_2 \) according to \( \gamma(dx_2|x_1) \) and thus \((3.9)\) denotes the total cost of transportation using this plan. When a probability measure \( \gamma \) is clearly specified
by the context, it will sometimes be denoted just by $P$, and the associated integral, or expectation, will be denoted by $E$. Couplings are very well-known in all branches of probability theory, we refer the reader to the second book of C. Villani [116] and L. Rüschendorf [103] for some basic reminders and a few more technical issues.

Of course, the solution of the Monge–Kantorovich problem depends on the cost function $c$. The cost function and the probability spaces here can be very general, and some nontrivial results can be obtained as soon as, say, $c$ is lower semicontinuous and $\mathcal{X}_i$, $i = 1, 2$ are Polish spaces. We observe that couplings always exist: at least there is the trivial coupling, in which the variables $X_i$, $i = 1, 2$ are independent (so their joint law is $\mu_1 \otimes \mu_2$). Under certain assumptions one can guarantee that the optimal coupling $(X_1, X_2)$ really is deterministic; the search of deterministic optimal “Monge” couplings is called the *Monge problem*. By “deterministic coupling” we mean there is a measurable function $T : \mathcal{X}_1 \to \mathcal{X}_2$ such that $X_2 = T(X_1)$, i.e. whenever $T$ is a change of variables from $\mu_1$ to $\mu_2$. A solution of the Monge problem yields a plan to transport the mass at minimal cost with a recipe that associates to each point $x_1$ a single point $x_2$ (“No mass shall be split”). Indeed, a subclass of all transport plans are deterministic transport plans of the form $\gamma(\cdot | x_1) = E_{T(x_1)}$, where $T$ is a map which transports $\mu_1$ to $\mu_2$; that is, $T_#\mu_1 = \mu_2$. The additional restriction is that no mass is allowed to be split.

**Proposition 3.7** Let $T(\mu_1, \mu_2)$ be the set of all deterministic transport plans $T : \mathcal{X}_1 \to \mathcal{X}_2$ such that $T$ is measurable and $T_#\mu_1 = \mu_2$. Then, one obtains the corresponding Monge transportation problem (MP), respectively the deterministic coupling problem

$$\inf_{T \in T(\mu_1, \mu_2)} E[c(X_1, T(X_1))], \quad (3.10)$$

where the pair $(X_1, T(X_1))$ runs over all possible couplings of $(\mu_1, \mu_2)$. In other words, the basic Monge–Kantorovich problem of optimal transport consists in the minimization problem

$$\inf \left\{ E[c(X_1, X_2)] : \text{law}(X_1) = \mu_1, \text{law}(X_2) = \mu_2 \right\}.$$

In this context $c$ is understood as a distance (dissimilarity) and (3.10) is a natural problem.
of “Monge–Kantorovich” optimization to find an coupling \((X_1, X_2)\) of \((\mu_1, \mu_2)\) with minimal expected dissimilarity. Such joint measures are called transference (or transport) plans; those achieving the infimum are called optimal transference plans.

**Corollary 3.8** The problem \((KP)\) is equivalent to find an optimal coupling of \((\mu_1, \mu_2)\) with respect to the coupling function \(c\) that satisfying \((3.10)\) over all couplings \((X_1, X_2)\) of \((\mu_1, \mu_2)\).

### 3.2.2 Existence of optimal transport plans and couplings

One might wonder now how generally are Monge’s and Kantorovich’s formulations the same in the sense that the minimums in the problems \((MP)\), \((KP)\) agree. It can be shown that they agree for example when the starting measure \(\mu_1\) has no atoms; that is, \(\mu_1(x_1) = 0\) for all \(x_1 \in X_1\), and the cost function \(c\) is continuous. As our first step, we give a very rough answer to the primal Monge–Kantorovich minimization problem. Trying to solve Monge problem, one has to face many problems: a first problem is that it may happen that there are no optimal transport maps, as we saw with Example 3.3 above. Furthermore, the lack of any notation of coercivity (in any Sobolev or Sobolev–like space) for the Monge functional (as there is no dependence on any gradients) makes the Monge’s problem unapproachable by (standard) methods of Calculus of Variations; however, this precisely the feature that opens Kantorovich’s problem to linear programming [58], [115].

If one is interested in the problem of Monge, a number of questions naturally arise at this point. Do there exist minimizes for the Kantorovich problem? If so, can they be shown to be minimizers of the Monge problem as well? Are there cases when Kantorovich’s problem can be solved but Monge’s problem cannot? What about vice-versa? To answer these questions let’s begin with the last question that may become “does this minimal \(\gamma\) come from a transport map \(T\)?”. Actually, if the answer to this question is yes, then it is evident that the problem of Monge has a solution, which also solves a wider problem, that of minimizing among transport plans. On the other hand, in some cases proving that the optimal transport plan comes from a transport map (or proving that there exists at least one optimal plan coming from a map) is equivalent to proving that the problem of Monge has a solution, since very often the infimum among transport plans and among transport maps
is the same. This depends on the presence of atoms. The following interesting example is included in the statement of Monge problem (MP) and where an optimal transport exists.

**Example 3.9** For any $b \in \mathcal{X}_2$, obviously, there exists a unique transfer map, namely $S(x_1) = b$ for all $x_1 \in \mathcal{X}_1$, pushing $\mu_1$ into $\delta_b$. Furthermore, by the procedure described above, this map yields the only transfer plan between $\mu_1$ and $\delta_b$. Therefore, the solution of the Monge problem coincides to that of the Kantorovich problem (KP) and the optimal cost (MP) is

$$ T_c(\mu_1, \delta_b) = \int_{\mathcal{X}_1} c(x_1, b) \, d\mu_1(x_1) $$

(We refer the reader to [115], [116] for more examples illustrating different situations.)

On the contrary, the problem (KP) by Kantorovitch still makes sense even for atomic measures, since we already said that there always exists a transport plan $\gamma$ between any two given probability measures. In particular, the situation is particularly easy as seen in Example 3.3 when $\mu_1 = \delta_0$ and $\mu_2$ is not a single Dirac mass. In that case the set $\Gamma(\mu_1, \mu_2)$ contains a unique element, which is $\gamma = \mu_1 \otimes \mu_2 = \delta_0 \otimes \mu_2$. This can be checked in this way: take $\gamma \in \Gamma(\mu_1, \mu_2)$ and a test function $\xi : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$. Let us integrate $\int \xi \, d\gamma$: if we consider $\gamma((\{0\} \times \mathcal{X}_2)^c) = \gamma(\{0\}^c \times \mathcal{X}_2) = \mu_1(\{0\}^c) = 0$, we deduce that is concentrated on $\{0\} \times \mathcal{X}_2$; that is, that $x_1 = 0 \gamma$-a.e. This allows to replace the variable $x_1$ with 0 in the integration, thus getting $\int \xi \, d\gamma = \int \xi(0, x_2) \, d\gamma(x_1, x_2) = \int \xi(0, x_2) \, d\mu_2(x_2) = \int \xi \, d\delta_0 \otimes \mu_2$, which, indeed, proves $\gamma = \mu_1 \otimes \mu_2$.

Notice that, it is not our intention to cover these issues later in any detail, mainly because our work does not focus on them. We will, however, discuss the optimal total costs $C(\mu_1, \mu_2)$ in some details, as they are the means that we make the most use of. Therefore, we will only state and briefly discuss some important results which are feel are necessary for a better understanding of what comes after them. Before going on, let’s recall the following definition:

**Definition 3.10** We recall that a cost function $c$ on $\mathcal{X}_1 \times \mathcal{X}_2$ is a lower semicontinuous if it is lower semicontinuous at every point of its domain; that is, at $(\bar{x}_1, \bar{x}_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ it satisfies
The following theorem has already been given in [115], [116] under very mild assumptions on the cost function and the space. Indeed, the same result of the existence of an optimal transport plan \( \gamma \) can be given if we have much stronger assumptions of the continuity on \( c \) and compactness on \( X_i, i = 1, 2 \). We will repeat the theorem briefly with an argument only uses the lower semicontinuity of the cost function. The proof in [116] relies on basic variational arguments involving the topology of weak convergence (that is, imposed by bounded continuous test functions). There are two key properties to check: (i) compactness, (ii) lower semicontinuity. These issues are taken care of respectively in Lemmas 4.3 and 4.4 [116], which will be used again in the sequel. On one hand, as \( \mu_i, i = 1, 2 \) are inner regular, the set \( \Gamma(\mu_1, \mu_2) \) is tight and thus, being obviously closed, compact according to Prokhorov's theorem (see Theorem A.1): if \( X \) is a Polish space, then a subset \( \mathcal{P} \) of \( \mathcal{P}(X) \) is precompact for the weak topology if and only if it is tight; that is, for any \( \varepsilon > 0 \) there is a compact set \( K \subset X \) such that \( \mu[X \setminus K] \leq \varepsilon \) for all \( \mu \in \mathcal{P} \). On the other hand, as \( c \) is lower semicontinuous, the map \( \gamma \mapsto \int c(x_1, x_2) d\gamma(x_1, x_2) \) is also lower semicontinuous.

**Theorem 3.11** Let \((X_i, \mu_i), i = 1, 2\) be two Polish probability spaces and the cost function \( c : X_1 \times X_2 \to \mathbb{R} \cup \{+\infty\} \) be a lower semicontinuous and bounded from below. Then Problem (KP) admits a solution. In other words, there exists a minimizer \( \gamma^* \) of the problem (KP); that is, an element in \( \Gamma(\mu_1, \mu_2) \) such that \( K_c(\gamma^*) = \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} K_c(\gamma) \), i.e. \( \gamma \in \Gamma_c(\mu_1, \mu_2) \) (the set of all such solutions—see Remark 3.4).

**Proof.** **Step 1.** First of all, as notice in Remark 3.4 before, \( \Gamma(\mu_1, \mu_2) \) is not empty. The key point consists in noting that \( \Gamma(\mu_1, \mu_2) \) is compact for the weak topology of probability measures. To prove that we need to use Prokhorov theorem. This means showing that any sequence in any sequence in \( \Gamma(\mu_1, \mu_2) \) is tight. Since \( X_i, i = 1, 2 \) are Polish spaces, then we have seen (by Corollary 1.16 [98]) that \( \{\mu_i\}, i = 1, 2 \) are tight sets. Therefore, by Proposition 1.13 [98], \( \Gamma(\mu_1, \mu_2) \) is a tight set in \( \mathcal{P}(X_1 \times X_2) \). To do that, fix \( \varepsilon > 0 \) and let \( K_i, i = 1, 2 \) be
two compact subsets of \( \mathcal{X}_i \), respectively; such that \( \mu_i(\mathcal{X}_i \setminus \mathcal{K}_i) < \varepsilon / 2 \) for \( i = 1, 2 \). Then the set \( \mathcal{K}_1 \times \mathcal{K}_2 \) is compact in \( \mathcal{X}_1 \times \mathcal{X}_2 \) and, for any \( \gamma \in \Gamma(\mu_1, \mu_2) \), we have

\[
\gamma \left[ (\mathcal{X}_1 \times \mathcal{X}_2) \setminus (\mathcal{K}_1 \times \mathcal{K}_2) \right] = \gamma \left[ ([\mathcal{X}_1 \setminus \mathcal{K}_1] \times \mathcal{X}_2) \cup (\mathcal{X}_1 \times [\mathcal{X}_2 \setminus \mathcal{K}_2]) \right] \\
\leq \gamma \left[ (\mathcal{X}_1 \setminus \mathcal{K}_1) \times \mathcal{X}_2 \right] + \gamma \left[ \mathcal{X}_1 \times (\mathcal{X}_2 \setminus \mathcal{K}_2) \right] \\
= \mu_1(\mathcal{X}_1 \setminus \mathcal{K}_1) + \mu_2(\mathcal{X}_2 \setminus \mathcal{K}_2) < \varepsilon,
\]

where the last equality follows from the projection property \( \pi_i^* \gamma = \mu_i, \ i = 1, 2 \). This gives the desired tightness. Thus, \( \Gamma(\mu_1, \mu_2) \) has compact closure (with respect to \( w^*\)-topology) in \( \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2) \).

We claim that actually \( \Gamma(\mu_1, \mu_2) \) is weakly closed, and therefore it is weakly compact. Indeed, let \( (\gamma_k)_{k \in \mathbb{N}} \subset \Gamma(\mu_1, \mu_2) \) be a minimizing sequence for the Kantorovich optimization problem (KP). Therefore, the sequence \( (\gamma_k)_{k \in \mathbb{N}} \) is tight in \( \Gamma(\mu_1, \mu_2) \), in the sense of Prokhorov theorem, Theorem 4.2.3 [24], and the Prokhorov compactness theorem allows to extract a subsequence \( (\gamma_{k_j})_{j \in \mathbb{N}} \) which converges narrowly to a probability \( \gamma \) on \( \mathcal{X}_1 \times \mathcal{X}_2 \) in the sense of Definition 4.2.2 [24], which is equivalent to \( \gamma \in \Gamma(\mu_1, \mu_2) \). To show that \( \gamma \) is a transport plan, that is, it satisfies the projection property: take a Borel set \( B_1 \) of \( \mathcal{X}_1 \); for every compact set \( \mathcal{K}_1 \subset B_1 \) and every open set \( \mathcal{O} \supset B_1 \), by using the Alexandrov proposition, Proposition 4.2.3 [24], and the projection property of \( \gamma_k \), we have

\[
\mu_1(\mathcal{K}_1) = \limsup_{k \to +\infty} \gamma_k(\mathcal{K}_1 \times \mathcal{X}_2) \leq \gamma(\mathcal{K}_1 \times \mathcal{X}_2) \leq \gamma(\mathcal{O} \times \mathcal{X}_2) \\
\leq \liminf_{k \to +\infty} \gamma_k(\mathcal{O} \times \mathcal{X}_2) = \mu_1(\mathcal{O}).
\]

Since \( \mathcal{K}_1 \) and \( \mathcal{O} \) were arbitrarily chosen, we have \( \gamma(B_1 \times \mathcal{X}_2) = \mu_1(B_1) \). In a similar way we obtain the second projection property \( \gamma(\mathcal{X}_1 \times B_2) = \mu_2(B_2) \). Therefore, if \( \gamma_k \xrightarrow{w^*} \gamma \), then \( \gamma \in \Gamma(\mu_1, \mu_2) \) and hence, \( \Gamma(\mu_1, \mu_2) \) is weakly closed as claimed.

**Step 2.** To conclude, it remains to show that the transport plan \( \gamma^* \) is optimal for the Kantorovich problem (KP). Let \( (\gamma_k)_{k \in \mathbb{N}} \) be a minimizing sequence and let \( \gamma^* \in \Gamma(\mu_1, \mu_2) \) a cluster point the sequence (it exists due to compactness of \( \Gamma(\mu_1, \mu_2) \)). Since \( c \) is lower semicontinuous and nonnegative, there exits a nondecreasing sequence \((c_n)_{n \in \mathbb{N}}\) of continuous
and bounded “Lipschitz” functions which converge to $c$ pointwise, i.e. $c$ is the pointwise supremum of the $c_n$. Indeed, it is enough to consider the Moreau–Yosida approximation of $c(x_1, x_2)$, that is

$$\tilde{c}_n(x_1, x_2) := \inf_{(\tilde{x}_1, \tilde{x}_2) \in X_1 \times X_2} \left\{ c(\tilde{x}_1, \tilde{x}_2) + n d((x_1, x_2), (\tilde{x}_1, \tilde{x}_2)) \right\},$$

and to make it bounded by setting $c_n(x_1, x_2) := \min\{\tilde{c}_n(x_1, x_2), n\} = \tilde{c}_n \land n$. Obviously, $c_n$ is continuous, $c_n(x_1, x_2) \leq c(x_1, x_2)$, and $c_n$ converges pointwise to $c$. Finally, using respectively the fact that $\gamma^*$ is a cluster point, $c_n \leq c$, $\gamma_k$ is a minimizing sequence and the narrow convergence of $\gamma_k$ to $\gamma$ for every $k \in \mathbb{N}$, we can conclude

$$\int c_n \, d\gamma^* = \lim_{k \to +\infty} \int c_n \, d\gamma_k \leq \liminf_{k \to +\infty} \int c \, d\gamma_k = \inf_{\gamma} K_c(\gamma).$$

Passing now to the limit as $n \to +\infty$, the monotone convergence theorem gives the conclusion and so $\gamma^*$ is actually a minimizer of (KP). Thus, any minimizing sequence converges, up to an extraction, to a minimizer.

Note that the lower bound assumption on $c$ guarantees that the expectation of $c(X_1, X_2) \geq \alpha_1(X_1) + \alpha_2(X_2)$ is well–defined in $\mathbb{R} \cup \{+\infty\}$.

**Remark 3.12** The existence of an optimal coupling is exactly the same as the theorem above. In addition to the underlying assumptions made in Theorem 3.11, if we assume that $\alpha_1 : \mathcal{X}_1 \to \mathbb{R} \cup \{-\infty\}$ and $\alpha_2 : \mathcal{X}_2 \to \mathbb{R} \cup \{-\infty\}$ are two upper semicontinuous functions such that $\alpha_1 \in L^1(\mu_1)$, $\alpha_2 \in L^1(\mu_2)$ and $c(x_1, x_2) \geq \alpha_1(x_1) + \alpha_2(x_2)$ for all $x_1, x_2$, then there is a coupling of $(\mu_1, \mu_2)$ which minimizes the total cost (3.10) among all possible couplings $(X_1, X_2)$ of $(\mu_1, \mu_2)$. 

3.2.3 Dual optimization problem

For a general cost function \( c \) in Polish probability spaces \((\mathcal{X}_i, \mu_i), i = 1, 2\), one can derive an intuition that leads to states Theorem 3.16 below. We are given a "measurable" cost function \( c : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathbb{R}^+ \cup \{+\infty\} \) which will be nonnegative and continuous in most of our applications. First, we set up the cost function in such a way that

\[
\xi_2(x_2) - \xi_1(x_1) \leq c(x_1, x_2) \quad \text{for every} \quad (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2.
\]

where \( \xi_i : \mathcal{X}_i \rightarrow \mathbb{R} \cup \{+\infty\}, i = 1, 2 \) are bounded and continuous functions. Since the problem (KP) is a linear optimization under linear constraints, an important tool will be duality theory, which is typically used for convex problems [86]. We will find a dual formulation (3.14) for the Kantorovich problem (KP) and exploit the relations between dual and primal problems. While the central notion in the original Monge–Kantorovich problem is cost, in the dual problem it is price. Therefore, we can equivalently define that, \( \gamma \in \Gamma(\mu_1, \mu_2) \) if and only if for every \( (\xi_1, \xi_2) \in C_b(\mathcal{X}_1 \times \mathcal{X}_2) \) we have

\[
\int_{\mathcal{X}_1 \times \mathcal{X}_2} [\xi_2(x_2) - \xi_1(x_1)] d\gamma(x_1, x_2) = \int_{\mathcal{X}_2} \xi_2(x_2) d\mu_2(x_2) - \int_{\mathcal{X}_1} \xi_1(x_1) d\mu_1(x_1).
\]

We define,

\[
K_c(\gamma) := \int_{\mathcal{X}_1 \times \mathcal{X}_2} c(x_1, x_2) d\gamma(x_1, x_2) \quad \text{for every} \quad \gamma \in \Gamma(\mu_1, \mu_2)
\]

and

\[
K(\xi_1, \xi_2) := \int_{\mathcal{X}_2} \xi_2(x_2) d\mu_2(x_2) - \int_{\mathcal{X}_1} \xi_1(x_1) d\mu_1(x_1) \quad \text{for every} \quad \xi_1 \in L^1(\mu_1), \xi_2 \in L^1(\mu_2).
\]

The first thing we will do is finding a dual problem, by means of an inf–sup exchange. Then, we express the constraint \( \gamma \in \Gamma(\mu_1, \mu_2) \) in the following way: if \( \gamma \in \mathcal{M}_+(\mathcal{X}_1 \times \mathcal{X}_2) \) (the set of a nonnegative measure on \( \mathcal{X}_1 \times \mathcal{X}_2 \)), then we have

\[
\sup_{\xi_1, \xi_2} \left\{ K(\xi_1, \xi_2) - \int_{\mathcal{X}_1 \times \mathcal{X}_2} [\xi_2(x_2) - \xi_1(x_1)] d\gamma(x_1, x_2) \right\} = \begin{cases} 0 & \text{if} \ \gamma \in \Gamma(\mu_1, \mu_2) \\ +\infty & \text{otherwise} \end{cases}
\]
where the supremum runs over all $\xi_1 \in C_0(\mathcal{X}_1)$ and $\xi_2 \in C_0(\mathcal{X}_2)$. Hence, one can remove the constraints on $\gamma \in \Gamma(\mu_1, \mu_2)$ if he adds the previous sup, since if they are satisfied nothing has been added, and if they are not one gets $+\infty$ and this will be avoided by the minimization. As will be shown in Theorem 3.16 on the following pages, there is a form of duality between the Monge–Kantorovich problem (MP)–(KP) and the following other problem:

**Problem 3.13 (Dual problem)** Given $\mu_i \in \mathcal{P}(\mathcal{X}_i), i = 1, 2$ and the cost function $c : \mathcal{X}_1 \times \mathcal{X}_2 \to [0, +\infty]$, we consider the problem

$$\max_{(\xi_1, \xi_2) \in \Phi_c} \left\{ K(\xi_1, \xi_2) := \int_{\mathcal{X}_2} \xi_2 \, d\mu_2 - \int_{\mathcal{X}_1} \xi_1 \, d\mu_1 \right\},$$

where $\Phi_c$ denotes the set of all pairs $(\xi_1, \xi_2) \in C_0(\mathcal{X}_1 \times \mathcal{X}_2)$ such that (3.11) holds. In other words, find $(\xi_1, \xi_2) \in \Phi_c$ such that $\int_{\mathcal{X}_2} \xi_2 \, d\mu_2 - \int_{\mathcal{X}_1} \xi_1 \, d\mu_1$ is maximal.

This is often called the dual or sometimes primal problem, because they are linked (see 3.16 below), and the space of signed Radon measures—where the Monge–Kantorovich problem is defined—is the dual of the space of continuous functions vanishing at infinity—where this new problem (3.12) is defined, even though the condition to vanish at infinity is irrelevant. Whatever the naming, the requirement $(\xi_1, \xi_2) \in C_0(\mathcal{X}_1 \times \mathcal{X}_2)$ can be relaxed to the condition $\xi_1 \in L^1(\mathcal{X}_1, \mu_1)$, $\xi_2 \in L^1(\mathcal{X}_2, \mu_2)$, so that the dual problem (DP) becomes satisfied.

It seems natural to look for a solution of the new problem (DP) among the pairs $(\xi_1, \xi_2)$ that saturate the condition, and therefore satisfy a Kantorovich potential (a function that realizing the maximum in (3.12)) and $c$–transform notations. We define $c$–convexity and $c$–concavity where convexity is replaced by the notation of $c$–concavity and the Legendre transform is replaced by the $c$–transforms.

**Definition 3.14 ($c$–convexity and $c$–concavity)** Let $\mathcal{X}_i, i = 1, 2$ be two nonempty sets and $c : \mathcal{X}_1 \times \mathcal{X}_2 \to (-\infty, +\infty]$. Then,

(i) A function $\xi_1 : \mathcal{X}_1 \to \mathbb{R} \cup \{+\infty\}$ is said to be a proper $c$–convex if there exists $\xi_2 : \mathcal{X}_2 \to \mathbb{R} \cup \{\pm \infty\}$, such that $\xi_1(x_1) = \sup_{x_2 \in \mathcal{X}_2}[\xi_2(x_2) - c(x_1, x_2)]$ for every $x_1 \in \mathcal{X}_1$. The domain of $\xi_1$ is, of course, $\text{Dom}(\xi_1) := \{x_1 \in \mathcal{X}_1 : \xi_1(x_1) < +\infty\} \neq \emptyset$. 


(ii) If $\xi_1$ is not necessarily a proper $c$-convex, then its $c$-transform is the function $\xi_1^c : \mathcal{X}_2 \to \mathbb{R} \cup \{-\infty\}$ defined by $\xi_1^c(x_2) = \inf_{x_1 \in \mathcal{X}_1} [\xi_1(x_1) + c(x_1, x_2)]$ for every $x_2 \in \mathcal{X}_2$. The functions $\xi_1$ and $\xi_1^c$ are said to be $c$-conjugate.

(iii) The $c$-subdifferential of $\xi_1$ is the is the $c$-cyclically monotone set defined by

$$\partial_c \xi_1 := \{ (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 ; \; \xi_1^c(x_2) - \xi_1(x_1) = c(x_1, x_2) \} .$$

Moreover, the $c$-subdifferential of $\xi_1$ at point $x_1 \in \mathcal{X}_1$; denoted by $\partial_c \xi_1(x_1)$, is the set of $x_2 \in \mathcal{X}_2$ such that $(x_1, x_2) \in \partial_c \xi_1$; or equivalently, for all $\tilde{x}_1 \in \mathcal{X}_1$ there holds $\xi_1(x_1) + c(x_1, x_2) \leq \xi_1(\tilde{x}_1) + c(\tilde{x}_1, x_2)$.

(iv) With the same notations as in (i)–(iii), a function $\xi_2 : \mathcal{X}_2 \to \mathbb{R} \cup \{-\infty\}$ is said to be a proper $c$-concave if there exists $\xi_1 : \mathcal{X}_1 \to \mathbb{R} \cup \{\pm \infty\}$, such that $\xi_2 = \xi_1^c$. Note here $\text{Dom}(\xi_2) := \{ x_2 \in \mathcal{X}_2 : \xi_2(x_2) > -\infty \} \neq \emptyset$ and $\xi_1 = \xi_1^{cc}$. Then its $c$-transform is defined by $\xi_2^c = \xi_1$, and its $c$-superdifferential is the $c$-cyclically monotone set defined by

$$\partial^c \xi_2 := \{ (x_1, x_2) : \xi_2(x_2) - c(x_1, x_2) \geq \xi_2(\tilde{x}_2) - c(x_1, \tilde{x}_2) \text{ for every } \tilde{x}_2 \in \mathcal{X}_2 \} .$$

If we have equality between the minimum of (KP) and the maximum of (DP) and both extremal values in (3.12) are realized, one can consider an optimal transport plan $\gamma$ and a Kantorovich potential $\xi_1$ (the solution of the dual problem 3.13) and write $\xi_1^c - \xi_1 \leq c$ on $\mathcal{X}_1 \times \mathcal{X}_2$, where $\xi_1$, by Definition 3.14(i), can be chosen as the $c$-transform of $\xi_2$ and that the optimal transport plan $\gamma$ is supported by the closed set (where it is concentrated)

$$\text{spt}(\gamma) := \{ (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 : \xi_2(x_2) - \xi_1(x_1) = c(x_1, x_2) \} .$$

(3.13)

In fact, the equality on $\text{spt}(\gamma)$ is a consequence of the inequality which is valid everywhere and of $K_c(\gamma) = K(\xi_1, \xi_1^c) = \int \int [\xi_1^c - \xi_1] \, d\gamma$, which implies equality $\gamma$–a.e. These functions being continuous, the equality passes to the support of the measure $\gamma$. Then, the problem (DP) becomes
Proposition 3.15 Suppose that $\mathcal{X}_i, i = 1, 2$ are compact and $c$ is continuous. Then, the problem (DP) admits a Kantorovich potential solution $\xi_1 \in L^1(\mu_1)$ between $\mu_1$ and $\mu_2$ such that $K(\xi_1, \xi_1^c) := \int_{\mathcal{X}_2} \xi_1 d\mu_2 - \int_{\mathcal{X}_1} \xi_1 d\mu_1$ is maximal.

For a proof of this proposition, see Proposition 1.11 of the monograph by Filippo Santambrogio [104]. Now, with this proposition and the convex analysis notations defined above, the classical Monge–Kantorovich problem (MP) can be solved by the following main theorem. Before we state the duality theorem, let us denote by $\Xi_c$ the set of all pairs $(\xi_1, \xi_2) \in L^1(\mu_1) \times L^1(\mu_2)$ such that (3.11) holds for $i = 1, 2$, $\mu_i$–a.e. $x_i \in \mathcal{X}_i$.

Theorem 3.16 (Kantorovich duality) Let $(\mathcal{X}_i, \mu_i), i = 1, 2$ be two Polish probability spaces and the cost function $c : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous and bounded from below. Then the following statements hold:

(i) There is duality which can be extended in the following way

\[
\inf_{\gamma \in \Gamma(\mu_1, \mu_2)} K_c(\gamma) = \sup_{\xi_1, \xi_2 \in \Xi_c} K(\xi_1, \xi_2) = \sup_{\xi_1 \in L^1(\mu_1)} K(\xi_1, \xi_1^c) = \sup_{\xi_2 \in L^1(\mu_2)} K(\xi_2^c, \xi_2), \quad (3.14)
\]

where $\xi_1^c, \xi_2^c$ is the $c$–transforms of $\xi_1, \xi_2$; respectively. Moreover, the infimum in the left–hand side of (3.14) is attained. Furthermore, the value of the first supremum in the right–hand side of (3.14) does not change if the set $\Xi_c$ is restricted to only pairs of bounded and continuous functions $C_b(\mathcal{X}_1 \times \mathcal{X}_2)$, and one might as well impose that $\xi_1$ is $c$–convex or $\xi_2$ is $c$–concave in the last two suprema.

(ii) If $c$ is real–valued, and the optimal cost $\mathcal{C}(\mu_1, \mu_2)$ is finite, then there exists a measurable and $c$–cyclically monotone set $\mathcal{G} \subset \mathcal{X}_1 \times \mathcal{X}_2$ (additionally, $\mathcal{G}$ is closed, if $c$ is continuous) such that for any $\gamma \in \Gamma(\mu_1, \mu_2)$ the following facts are equivalent:

(a) $\gamma$ is optimal;

(b) $\gamma$ is $c$–cyclically monotone ($\text{spt}(\gamma)$ is $c$–cyclically monotone when $c$ is continuous);

(c) There is a $c$–convex $\xi_1$ such that, $\gamma$–almost surely, $\xi_1^c(x_2) - \xi_1(x_1) = c(x_1, x_2)$;

(d) There exist $\xi_1 : \mathcal{X}_1 \to \mathbb{R} \cup \{+\infty\}$ and $\xi_2 : \mathcal{X}_2 \to \mathbb{R} \cup \{-\infty\}$, such that $\xi_2(x_2) - \xi_1(x_1) \leq c(x_1, x_2)$ for all $(x_1, x_2)$, with equality $\gamma$–almost surely;
(e) $\gamma$ is concentrated on $G$ ($\text{spt}(\gamma) \subset G$ when $c$ is continuous).

(iii) If $c$ is real–valued, and $C(\mu_1, \mu_2) < +\infty$, and one has the pointwise upper bound

$$c(x_1, x_2) \leq c_{\mathcal{X}_1}(x_1) + c_{\mathcal{X}_2}(x_2), \quad (c_{\mathcal{X}_1}, c_{\mathcal{X}_2}) \in L^1(\mu_1) \times L^1(\mu_2), \quad (3.15)$$

then both the primal and dual Kantorovich problems have solutions, so

$$\min_{\gamma \in \Gamma(\mu_1, \mu_2)} K_c(\gamma) = \max_{\xi_1, \xi_2 \in \Xi_c} K(\xi_1, \xi_2) = \max_{\xi_1 \in L^1(\mu_1)} K(\xi_1, \xi_1^c),$$

and in the latter expressions one might as well impose that $\xi_1$ be $c$–convex and $\xi_2 = \xi_1^c$. If in addition $c$ is continuous, then there is a closed $c$–cyclically monotone set $G \subset X_1 \times X_2$, such that for any $\gamma \in \Gamma(\mu_1, \mu_2)$ and for any $c$–convex $\xi_1 \in L^1(\mu_1)$,

$$\begin{cases} 
\gamma \text{ is optimal in the Kantorovich problem if and only if } \gamma(G) = 1; \\
\xi_1 \text{ is optimal in the dual Kantorovich problem if and only if } G \subset \partial_c \xi_1.
\end{cases}$$

We will not prove Theorem 3.16. We refer to [116], Theorem 5.10, for a complete proof, with slightly different notations. This theorem conveys a very clear picture of the relationship between primal and dual solutions, and, in particular, a cyclically monotone transference plan leads to the conclusion that all transference plans should be cyclically monotone. If $\gamma \in \Gamma(\mu_1, \mu_2)$ is an optimal transport plan; that is, a solution of (KP), the theorem shows that $\gamma$ is necessarily induced by a transport map; that is, is of the form $\gamma = \gamma_T$ for some transport $T$. The main tool is the duality result; that is, $(\xi_1, \xi_2)$ is an optimal solution to the dual problem as introduced by (3.12).

**Remark 3.17** (i) The first part of Theorem 3.16 asserts that solving the primal Kantorovich problem (KP) is equivalent to solve the duality formula (3.12). Something similar to the first part will be recovered also in the setting of Theorem 3.19. Indeed, we will show that a similar “restriction” can be performed for a slightly modified Kantorovich duality, obtained exploiting (3.18) and the quadratic cost $c(x_1, x_2) = |x_1 - x_2|^2$ on $\mathbb{R}^d$. Namely, we might consider only couples $(\xi_1, \xi_1^c)$ with $\xi_1$ convex and lower semicontinuous (see Villani first book [115]).
(ii) The second part of Theorem 3.16, states the equivalence between the properties of optimality (of a transference plan $\gamma$) and $c$-cyclical monotonicity. This is possible because we added the hypothesis of a real valued cost $c$, for quite general cost functions and probability measures, the problem was completely solved recently (A possible construction of an equivalent condition to optimality is sketched in [116], but it requires a lot of measure theory and it is far beyond our goal here). The current state of the art is:

(a) the equivalence is true for a real-valued lower semicontinuous cost function $c$ (that is, Theorem 3.16);

(b) the equivalence is true for a continuous cost function $c$ with possibly attaining infinite values (see [97]);

(c) the equivalence is false for a general lower semicontinuous cost function $c$ with possibly infinite values (see [19] for a counterexample).

(iii) Since the existence theorem, Theorem 3.11 does not imply that the optimal cost is finite. It might be that all transport plans lead to an infinite total cost; that is, $K_c(\gamma) = +\infty$ for all $\gamma \in \Gamma(\mu_1, \mu_2)$. To exclude this possibility, we assumed in Theorem 3.16(iii) a simple condition (3.15), which implies that any coupling has finite total cost, and has other nice consequences. The assumption (3.15) can be weakened into

\[
\int_{X_1 \times X_2} c(x_1, x_2) \, d\mu_1 \otimes \mu_2(x_1, x_2) < +\infty \tag{3.16}
\]

which guarantees that at least the independent coupling has finite total cost, or even

\[
\mu_1\{x_1 : \int_{X_2} c(x_1, x_2) \, d\mu_2(x_2) < +\infty\} > 0, \quad \mu_2\{x_2 : \int_{X_1} c(x_1, x_2) \, d\mu_1(x_1) < +\infty\} > 0,
\]

which are implied by (3.16). These conditions are sufficient to prove an optimality for the Monge–Kantorovich minimization problem. We refer the reader to [17] for more details.

Notice that, we have assumed $c \geq 0$. All the results in Theorem 3.16 still hold true and could be immediately extended to the more general case of a cost function satisfying a lower bound as the following one

\[
c(x_1, x_2) \geq \alpha_1(x_1) + \alpha_2(x_2) \quad \text{for every} \quad (x_1, x_2) \in X_1 \times X_2, \tag{3.17}
\]
for some real-valued upper semicontinuous functions $\alpha_1 \in L^1(\mu_1)$ and $\alpha_2 \in L^1(\mu_2)$. Nevertheless, dealing with $c$-convexity and $c$-concavity, one has to slightly modify the following argument. Indeed, we can always define

$$\tilde{c}(x_1, x_2) := c(x_1, x_2) - \alpha_1(x_1) - \alpha_2(x_2) \geq 0, \quad \Lambda := \int_{X_1} \alpha_1 \, d\mu_1 + \int_{X_2} \alpha_2 \, d\mu_2 \in \mathbb{R}.$$ 

Whenever $\xi_1 : X_1 \to \mathbb{R} \cup \{+\infty\}$ and $\xi_2 : X_2 \to \mathbb{R} \cup \{-\infty\}$ are two functions, define

$$\tilde{\xi}_1(x_1) := \xi_1(x_1) + \alpha_1(x_1), \quad \tilde{\xi}_2(x_2) := \xi_2(x_2) - \alpha_2(x_2).$$

With these definitions, the following properties can be checked easily:

1. If $c$ is a real-valued lower semicontinuous cost function, then so does $\tilde{c}$;
2. $(\tilde{\xi}_1, \tilde{\xi}_2) \in L^1(\mu_1) \times L^1(\mu_2)$ if and only if $(\xi_1, \xi_2) \in L^1(\mu_1) \times L^1(\mu_2)$;
3. $K_{\tilde{c}}(\gamma) = K_c(\gamma) - \Lambda$ and $K(\tilde{\xi}_1, \tilde{\xi}_2) = K(\xi_1, \xi_2) - \Lambda$ for every $\gamma \in \Gamma(\mu_1, \mu_2)$ and $(\xi_1, \xi_2) \in \Xi_c$.
4. $\xi_1$ (resp. $\xi_2$) is $c$-convex (resp. $c$-concave) if and only if $\tilde{\xi}_1$ (resp. $\tilde{\xi}_2$) is $\tilde{c}$-convex (resp. $\tilde{c}$-concave).
5. $(\xi_1, \xi_2)$ is $c$-conjugate if and only if $(\tilde{\xi}_1, \tilde{\xi}_2)$ is $\tilde{c}$-conjugate, where $(\xi_1, \xi_2)$ are said to be $c$-conjugate if either $\xi_2 = \xi_1^c$ or $\xi = \xi_2^c$.
6. $G \subset X_1 \times X_2$ is $c$-cyclically monotone if and only if $G$ is $\tilde{c}$-cyclically monotone.

Due to these statements of duality properties, it is equivalent to establish Theorem 3.16 for the cost $c$ or for the nonnegative cost $\tilde{c}$. Therefore, to any solution of the Kantorovich transportation problem (in both of its formulations) with nonnegative cost function there corresponds a solution of the Kantorovich transportation problem (as in (3)) with cost bounded below as in (3.17).

Finally, the Monge–Kantorovich optimization problem and some variants have been developed in the probabilistic literature since the mid seventies. For various kinds of optimization problems they have been established as a basic and natural tool. A detailed exposition with many applications of these developments is given in Rachev and Rüschendorf [101], [102]. From the probabilistic point of view the Kantorovich formulation of the transport problem is more “natural” than the Monge formulation. Similar extensions of deterministic optimization problems are quite often to find in probability and statistics as e.g. the
transition from deterministic decision rules (like deterministic tests and estimators) to randomized decision rules (like randomized tests and estimators) is a classical example from the early period of statistics. Kantorovich obviously was not aware of the Monge problem when he formulated his transport problem in 1942. In 1948 he wrote a short note of three pages where he made the connection to the Monge problem and stated that in case (MP) has a deterministic solution $T$, then $T$ is also a solution of the Monge problem; that is, the Kantorovich problem is a relaxation of the Monge problem. In fact Kantorovich’s problem was one of the earliest infinite dimensional linear programming problems considered. In 1975 Kantorovich got together with Koopman the Nobel price in economics for his development of linear programming and the application to mathematical economics.

Villani presented in his second book [116] different examples of the most famous couplings used in mathematics. One of them is classical and well–representative in the topics that will be considered later in this thesis, the so–called \textit{an increasing rearrangement on the real line}. We will discuss this coupling with more details in next chapters. Starting with the late eighties, early nineties, important connections of optimal couplings with problems arise in economics, physics, partial differential equations, analysis, geometry, fluid mechanics and many others have been detected (by the way, the increasing rearrangement can be seen as particular case of optimal transport). Optimal couplings are quite stable with respect to perturbations. They exist in smooth as well as nonsmooth settings. They come with a rich structure: an \textit{optimal cost} functional (the value of the infimum defining the Monge–Kantorovich problem); a \textit{dual variational problem}; and, under adequate structure conditions, a continuous \textit{interpolation}. This lead to a very active and wide ranged research area. This line of research is excellently described and developed in the books of Ambrosio [4], Ambrosio, Gigli, and Savaré [17], and Villani [115], [116].

### 3.3 Regularity of Brenier’s map and its applications

We summarize here some different and useful results due to Y. Brenier [34] induced for the general theory of optimal transportation, with a particular focus on the case of the \textit{quadratic cost} function defined on the multidimensional space. The cases of the general
displacement ground cost function \( c(x_1, x_2) = |x_1 - x_2|^p \) with \( p \geq 1 \) have been particularly studied in \( \mathbb{R}^d \), where the minimal transport cost \( C(\mu_1, \mu_2) \) of (KP) provides, through the relation \( W_p(\mu_1, \mu_2) = (C(\mu_1, \mu_2))^{1/p} \), the so–called Wasserstein distance (this case is addressed in next chapters for the unidimensional space). This distance metrizes the weak convergence on the space \( \mathcal{P}(\mathbb{R}^d) \) of probabilities on \( \mathbb{R}^d \). A very wide literature on the subject is available; we mention, for instance, the books [17], [104], [115], [116], where one can find a complete list of references. There are similar results due to R. McCann and W. Gangbo [69] for more general cost functions of the form \( c(x_1 - x_2) \) with \( c \) is strictly convex and smooth in a \( d \)-dimensional Euclidean space. All these cases allow for very strong results, and in particular we can find existence (and uniqueness), as well as a representation formulas, for the optimal map \( T \).

3.3.1 Existence and uniqueness of Brenier’s maps

The way to understand the intuition behind the very important result in the Kantorovich duality theorem above, Theorem 3.16, is by analogy with the finite dimensional situation. It is well–known, and widely used, that a linear minimization problem with convex constraints, like (KP), admits a dual formulation on \( \mathbb{R}^d \) for \( d \geq 1 \). Suppose \( c_{ij}, \alpha_i, \beta_i \) \((i, j = 1, \ldots, d)\) are nonnegative real number satisfying \( \sum_{i=1}^{d} \alpha_i = \sum_{i=1}^{d} \beta_i \) and we look for \( t_{ij} \) that minimize \( \sum_{i,j=1}^{d} c_{ij} s_{ij} \) subject to the constraints \( \sum_{j=1}^{d} s_{ij} = \alpha_i, \sum_{i=1}^{d} s_{ij} = \beta_j, s_{ij} \geq 0 \) for \( i, j = 1, \ldots, d \). There is a linear programming problem which is the dual of the one described in Theorem 3.16 and consistent of finding \( x_1, x_2 \in \mathbb{R}^d \) so as to maximize \( \sum_{i}^{d} (\alpha_i x_1^i + \beta_i x_2^i) \) subject to the constraint \( x_1^i + x_2^i \leq c_{ij} \) for \( i, j = 1, \ldots, d \).

The same kind of intuition mentioned above in the case of the quadratic cost \( c(x_1, x_2) = |x_1 - x_2|^2 \); that is, referring to the finite dimensional case or considering measures supported on a finite number of disjoint balls of equal volume, led to the realization that if an optimal transfer map \( T \) for the Monge problem exists, then the graph \( \{(x_1, T(x_1)) : x_1 \in \mathcal{X}_1\} \) is cyclically monotone [88], [58]. This means that for any positive integer \( d \) and any finite
sequence \( \{x_i^j\}_{i=1}^d \subset X_1 \), we have
\[
\sum_{i=1}^d x_i^j \cdot (T(x_i^{j+1}) - T(x_i^j)) \leq 0 \quad \text{where} \quad x_i^{j+1} := x_i^j.
\]

Obviously, here we have assumed \( X_i, i = 1, 2 \) to be the supports of \( \mu_i \); respectively, and subsets of \( \mathbb{R}^d \). There is a theorem, due to Rockafellar, asserting that a cyclically monotone graph lies in the subdifferential of convex map of \( \mathbb{R}^d \) into \( \mathbb{R} \). As convex functions are differentiable a.e., one gets the intuition that the optimal map is the gradient of a convex function.

**Remark 3.18** For general measures \( \mu_i, i = 1, 2 \) one cannot expect any uniqueness for the solutions of the Kantorovich Problem (KP). Indeed, consider \( \mu_i \in \mathcal{P}(\mathbb{R}^2), i = 1, 2 \) such that \( \mu_1 \) is concentrated in \( \{(0,0), (1,1)\} \) and \( \mu_2 \) is concentrated in \( \{(1,0), (0,1)\} \). In this situation it is clear that both the plans

\[
\gamma_1 : \begin{cases} 
(0,0) \mapsto (1,0), \\
(1,1) \mapsto (0,1),
\end{cases} \quad \gamma_2 : \begin{cases} 
(0,0) \mapsto (0,1), \\
(1,1) \mapsto (1,0),
\end{cases}
\]

are optimal for the quadratic distance cost \( c(x, x_2) = |x - x_2|^2 \). A figure of these two optimal transportation plans could help to visualize the situation.

We now state what [115] calls the optimal transportation theorem for quadratic cost. The first part is Knott–Smith optimality criterion for Kantorovich’s problem and the second, third and fourth parts are Brenier’s theorem for Monge’s problem.

**Theorem 3.19** Let \( \mu_i, i = 1, 2 \) be a probability measures on \( \mathbb{R}^d \) with a finite sum of the second order moments; that is,
\[
\Lambda_2 := \int_{\mathbb{R}^d} \sum_{i=1}^2 \frac{|x_i|^2}{2} d\mu_i(x_i) < +\infty. \tag{3.18}
\]

We consider the Monge–Kantorovich problem (KP) associated with a quadratic cost function \( c(x_1, x_2) = |x_1 - x_2|^2 \). Then the following facts hold:
(i) A plan $\gamma \in \Gamma(\mu_1, \mu_2)$ is optimal if and only if there exists a proper, convex and lower semicontinuous function $\xi_1 : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that $\text{spt}(\gamma) \subseteq \text{Graph}(\partial \xi_1)$, or, equivalently, $x_2 \in \partial \xi_1(x_1)$ for $\gamma$-a.e. $(x_1, x_2)$. Moreover, for every $\xi_i \in C_b(\mathbb{R}^d), i = 1, 2$, the couple $(\xi_1, \xi^*_1)$ is the minimizer that solves the dual Monge–Kantorovich problem

$$\inf \left\{ K(\xi_1, \xi_2) : x_1 \cdot x_2 \leq \xi_2(x_2) - \xi_1(x_1) \text{ for every } x_1, x_2 \in \mathbb{R}^d \right\}, \quad (3.19)$$

(where $\xi_1^*$ denotes the Legendre transform of $\xi_1$).

(ii) If $\mu_1$ does not give mass to sets with finite $\mathcal{H}^{d-1}$ (that is, $\mu_1(A) = 0$ if $A$ is of Hausdorff dimension at most $d-1$), then there exists a unique optimal transportation plan $\gamma$ and it is given by $\gamma = (\text{Id} \times \nabla \xi_1)_{\#} \mu_1$ (or equivalently, $d\mu_1(x_1) \otimes \delta_{x_2 = \nabla \xi_1(x_1)}$), where $\nabla \xi_1$ is the uniquely determined $\mu_1$-a.e. gradient of a convex function $\xi_1$ such that $\nabla \xi_1_{\#} \mu_1 = \mu_2$. Moreover, $\text{spt}(\mu_2) = \nabla \xi_1(\text{spt}(\mu_1))$.

(iii) As a consequence, $\nabla \xi_1$ is the unique solution to Monge's transportation problem (MP); that is, $\nabla \xi_1$ is the optimal transport map solving the following minimization problem

$$\inf_{T_{\#} \mu_1 = \mu_2} \int_{\mathbb{R}^d} |x_1 - T(x_1)|^2 d\mu_1(x_1). \quad (3.20)$$

(iv) If $\mu_2$ does not give mass to small sets (that is, sets with finite $\mathcal{H}^{d-1}$), then, for $\mu_i, i = 1, 2$–a.e. $x_1$, respectively, we have $\nabla \xi^*_1 \circ \nabla \xi_1(x_1) = x_1$ and $\nabla \xi_1 \circ \nabla \xi^*_1(x_2) = x_2$, where $\nabla \xi^*_1$ is the uniquely determined $\mu_2$-a.e. gradient of a convex function which pushes $\mu_2$ forward to $\mu_1$. Furthermore, $\nabla \xi^*_1$ is the unique solution of the Monge problem of transporting $\mu_2$ onto $\mu_1$ with a quadratic cost function.

Sketch of Proof. For an explicit proof of this theorem we refer to Theorem 2.12 [115]. Since we know by Theorem 3.11 that the existence of a minimiser for the Kantorovich problem (KP), in the following we only recover the ideas that reformulate the problem (3.14) exploiting (3.18) and the quadratic cost $\frac{1}{2}|x_1 - x_2|^2$ to give a rough idea of a possible strategy of proof.
Notice that a specific approach for the case $\frac{1}{2}|x_1 - x_2|^2$, based on the fact that we can withdraw the parts of the cost depending on $x_1$ or $x_2$ only and maximize $\int x_1 \cdot x_2 \, d\gamma$, gives the same result in a easier way: we actually get $\xi_1(\tilde{x}_1) + \xi_2^*(\tilde{x}_2) = \tilde{x}_1 \cdot \tilde{x}_2$ for a convex function $\xi_1$ and its Legendre transform $\xi_2^*$ and we deduce $\tilde{x}_2 \in \partial \xi(\tilde{x}_1)$. Indeed, we can rewrite (KP) as follows

$$
\inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|x_1 - x_2|^2}{2} \, d\gamma(x_1, x_2) = \Lambda_2 - \sup_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} x_1 \cdot x_2 \, d\gamma(x_1, x_2). \tag{3.21}
$$

On the other hand, we also have $(\xi_1, \xi_2) \in \Xi_c$ if and only if $\xi_2(x_2) - \xi_1(x_1) \leq \frac{1}{2}|x_1 - x_2|^2$ for $i = 1, 2$, $\mu_i$-a.e. $x_i$, respectively. Equivalently, $x_1 \cdot x_2 \leq (\frac{1}{2}|x_1|^2 + \xi_1(x_1)) + (\frac{1}{2}|x_2|^2 - \xi_2(x_2))$ for $i = 1, 2$, $\mu_i$-a.e. $x_i$, respectively. We find, after rearranging terms, $x_1 \cdot x_2 \leq \tilde{\xi}_i(x_2) - \tilde{\xi}_i(x_1)$ for $i = 1, 2$, $\mu_i$-a.e. $x_i$, respectively, where $\tilde{\xi}_i(x_1) := -\frac{1}{2}|x_1|^2 - \xi_1(x_1)$ and $\tilde{\xi}_2(x_2) := \frac{1}{2}|x_2|^2 - \xi_2(x_2)$ are new unknowns functions. Therefore, there holds

$$
\sup_{(\xi_1, \xi_2) \in \Xi_c} K(\xi_1, \xi_2) = \Lambda_2 - \inf_{(\xi_1, \xi_2) \in \Xi_c} K(\tilde{\xi}_1, \tilde{\xi}_2), \tag{3.22}
$$

where $\Xi_c$ the set of all pairs $(\tilde{\xi}_1, \tilde{\xi}_2) \in L^1(\mu_1) \times L^1(\mu_2)$ such that $x_1 \cdot x_2 \leq \tilde{\xi}_2(x_2) - \tilde{\xi}_1(x_1)$ for $i = 1, 2$, $\mu_i$-a.e. $x_i \in \mathcal{X}_i$; respectively. In the sequel, for notational convenience we shall forget about the $\tilde{\xi}$ symbol. Finally, from (3.14), (3.21) and (3.22) we can rewrite the Kantorovich duality theorem as follows

$$
\sup_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} x_1 \cdot x_2 \, d\gamma(x_1, x_2) = \inf_{(\xi_1, \xi_2) \in \Xi_c} K(\xi_1, \xi_2). \tag{3.23}
$$

Note that restricting our attention to the smaller set (3.19) in the infimum (3.23) does not make any difference (see [115] with a complete proof).

**Remark 3.20** Notice that $(\xi_1, \xi_2)$ does not solve the dual Monge–Kantorovich problem in Theorem 3.19(i). It is the couple $(\tilde{\xi}_1(x_1), \tilde{\xi}_2(x_2)) := (\frac{1}{2}|x_1|^2 + \xi_1(x_1), \frac{1}{2}|x_2|^2 - \xi_2(x_2))$, which provides a solution to the dual problem.

Let us moreover notice that there may be no measurable $T$ such that $T \# \mu_1 = \mu_2$. Thus it is natural to enforce some “regularity” assumption, as in a part (ii) of Theorem 3.19 above.
The assumption that $\mu_1$ gives no mass to small sets is crucial in some sense optimal, as shown by some Examples in [115].

Now, what about more general situations? A complete characterization, as in Therem 3.19, is not known at this point. Anyway, one could prove the following results under necessary and sufficient optimality conditions. First, without assuming finiteness of the second moments (3.18), the following theorem holds true.

**Theorem 3.21** Let assume the Monge–Kantorovich problem (KP) associated with a quadratic cost function $c(x_1, x_2) = |x_1 - x_2|^2$ on $\mathbb{R}^d$. Then the following facts of necessity and sufficiency conditions hold:

(i) If a plan $\gamma \in \Gamma(\mu_1, \mu_2)$ is optimal, then there exists $\xi_1 : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ proper, convex and lower semicontinuous function such that $\text{spt}(\gamma) \subset \text{Graph}(\partial \xi_1)$.

(ii) If $\mu_1$ does not give mass to sets with finite $\mathcal{H}^{d-1}$, then there exists a unique $\xi_1 : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ proper, convex and lower semicontinuous function such that $\nabla \xi_1 \# \mu_1 = \mu_2$

A complete proof of this theorem can be found in [88]. We refer also to [115] for a partial proof. The part (ii) is known as a refined version of Brenier’s Theorem.

**Remark 3.22** Comparing Theorem 3.21 to Theorem 3.19, we can see that here there is no more a characterization of optimality; that is, there is no proof that, for general measures $(\mu_1, \mu_2)$ transportation plans $\gamma$ concentrated on a cyclically monotone set are optimal. The equivalence holds when $\mu_1$ does not give mass to sets with $\mathcal{H}^{d-1}$ measure, and it is a consequence of the uniqueness in Brenier’s Theorem. It is also true when we assume the weakened sufficient condition (3.16); as proved in [19] for the case of strictly convex cost function, and in [116] for the general cost function. But these are the closest results known at present.

Let us now go back and have a closer look at the case of the quadratic cost which was initially solved by Y. Brenier in his groundbreaking article [34]; that is, let us consider (3.20) in Theorem 3.19(ii) above. We already know that there is a unique optimal transport $T$ that is characterized by the fact that $\xi_2(T(x_1)) - \xi_1(x_1) = \frac{1}{2} |x_1 - T(x_1)|^2$, where $\xi_1$ and $\xi_2$ are related by the conjugacy relations given by Definition 3.14 which can be rewritten due to
Remark 3.20 as $\tilde{\xi}_1(x_1) = \sup_{x_2} \{ \xi_2(x_2) + x_1 \cdot x_2 \}$ and $\tilde{\xi}_2(x_2) = \sup_{x_1} \{ x_1 \cdot x_2 - \tilde{\xi}_1(x_1) \}$, where the functions $\tilde{\xi}_1 := \frac{1}{2} | \cdot |^2 + \xi_1$ and $\tilde{\xi}_2 := \frac{1}{2} | \cdot |^2 - \xi_2$ are convex and conjugate to each other in the usual sense of convex analysis, i.e. $\xi_1 = \xi_2^*$ and $\xi_2 = \xi_1^*$. Therefore, Optimal transportation theorems for a strictly convex (resp. concave) cost function of the form $c(x_1 - x_2)$ (resp. $c(|x_1 - x_2|)$), which are given by Gangbo and McCann [68] (see also Theorems 2.44 and 2.45 of [115]), may be particularized to the quadratic case $c(x_1, x_2) = \frac{1}{2} |x_1 - x_2|^2$, thus getting the existence of an optimal transport map $T$ by the following formula

$$T(x_1) = x_1 - \nabla \xi_1(x_1) = \nabla \left( \frac{|x_1|^2}{2} - \xi_1(x_1) \right) = \nabla \tilde{\xi}_1(x_1)$$

for a convex function $\tilde{\xi}_1$. By using the converse implication (sufficient optimality conditions), this also proves the existence and uniqueness $\mu_1$ onto $\mu_2$. Indeed, Brenier’s theorem states that there exists a unique (up to $\mu_1$ negligible sets) map of the form $T(x_1) = \nabla \tilde{\xi}_1(x_1)$ with $\tilde{\xi}_1$ convex that transports $\mu_1$ to $\mu_2$, this map is also the optimal transport between $\mu_1$ and $\mu_2$ for the quadratic cost. In other words, the optimal transport $T$ is the gradient of a convex function $\xi_1$ and it is actually a characterization of optimality. This optimal map is called the Brenier’s map between $\mu_1$ and $\mu_2$.

**Corollary 3.23** Given $\mu_i \in \mathcal{P}(X_i)$, $i = 1, 2$ satisfying the condition (3.18). Then, if $\mu_1$ is absolutely continuous, there is a unique map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $\mu_2 = T_{\#} \mu_1$ and

$$\min_{T_{\#} \mu_1 = \mu_2} \int_{\mathbb{R}^d} |x_1 - T(x_1)|^2 d\mu_1(x_1) = \min_{\gamma \in T(\mu_1, \mu_2)} \int_{\mathbb{R}^d} |x_1 - x_2|^2 d\gamma(x_1, x_2).$$

Moreover, there is only one optimal transport plan $\gamma$, which is thus necessarily $(\text{Id}, T)_{\#} \mu_1$, and $T$ is the gradient of a convex function $\tilde{\xi}_1$, which is therefore unique up to an additive constant. There is also a unique (up to an additive constant) Kantorovich potential $\xi_1$, which is locally Lipschitz and linked to $\tilde{\xi}_1$ through the relation $\tilde{\xi}_1(x_1) = \frac{1}{2} |x_1|^2 - \xi_1(x_1)$.

As we mentioned before, for notational convenience we shall forget about the $\tilde{\cdot}$ symbol and consider the Brenier’s map $T = \nabla \xi$ in the following.
3.3.2 Monge–Ampère equation and regularity

Once the existence of Brenier’s map has been established, a natural question is about its regularity. Informally, the question can be stated as follows: *Given two smooth densities supported on good sets, it is true the optimal transport map is smooth?* Or, somehow more precisely, one can investigate how much is the “gain” in regularity from the densities to the optimal map. As we will see in a moment, a natural guess is that the optimal map should have “one derivative” more than densities.

A deep regularity theory due to L. Caffarelli [38], [37], and L. Evans [58] establishes conditions under which the optimal transport map is in fact smooth. If the function $\xi_1$ is smooth and strictly convex, and $\rho_i, i = 1, 2$ are the respective two densities of the absolutely continuous measures $\mu_i, i = 1, 2$, then we can write the condition $T\#\mu_1 = \mu_2$ in a PDE form known by the so–called *Monge–Ampère equation*

$$\det(\Delta \xi(x_1)) \rho_2(\nabla \xi(x_1)) = \rho_1(x_1), \quad (3.24)$$

This means that whenever $\rho_2$ is smooth enough, then $\xi$ must solve a nonlinear PDE (3.24) of a (degenerate) elliptic type. Clearly, the Monge–Ampère equation (3.24) can be interpreted as providing a weak (or generalized) solution to mean $\int \zeta(\nabla \xi(x_1)) \, d\rho_1(x_1) = \int \zeta(x_2) \, d\rho_2(x_2)$ for every continuous function $\zeta$.

Note that, we impose $\xi_1$ to be strictly convex just to guarantee injectivity of $T$. Indeed, the constraint $T\#\mu_1 = \mu_2$ is equivalent, if $\mu_i = \rho_i(x_i) \, dx_i, i = 1, 2$ and if $T$ is $C^1$ and injective, to the Jacobian condition

$$|\det(\nabla T(x_1))| = \frac{\rho_1(x_1)}{\rho_2(T(x_1))} \quad (3.25)$$

which can be obtained as a consequence of Proposition 1.11. It turns out to be a very degenerate first order PDE. As we already said, the above equation could lead to the guess that $T$ has one derivative more than the densities. Notice however that the above equation is satisfied by every map which satisfies

$$\int_{T^{-1}(B)} \rho_1(x_1) \, dx_1 = \int_B \rho_2(x_2) \, dx_2 \quad \text{for every Borel set } B \subset \mathbb{R}^d. \quad (3.26)$$
Thus, by simple examples, we cannot expect solutions of (3.25) to be well-behaved. Indeed, consider for instance the case in which $\rho_i = 1_{B_i}, \ i = 1, 2$ with $B_i$ smooth open sets. If we consider right (respectively left) compose $T_1$ with a map $T_2$ satisfying $\det \nabla T_2 = 1$ and $T_2(B_i) = B_i$ respectively for $i = 1, 2$, we still obtain a solution of (3.25) which is no more regular than $T_2$.

In the particular case when $T = \nabla \xi$, if $\mu_i = \rho_i \mathcal{L}^d, \ i = 1, 2$, then the optimal transport map $\nabla \xi$ from $\mathbb{R}^d$ into $\mathbb{R}^d$, at least formally satisfies the Monge–Ampère equation above as for boundary condition, is the requirement that $\nabla \xi$ maps the support of $\rho_1$ onto that of $\rho_2$. In this way there is hope to obtain regularity of $T = \nabla \xi$ from the regularity of the densities because; in fact, the equation (3.24) it is not enough to ensure regularity of $\xi$. A simple example is given by the case in which the support of the first density is connected while the support of the second is not (we refer, for instance, to [72], [96] for the general theory about this equation). Indeed, since by (3.26) it follows easily that $\nabla spt \rho_1 = spt \rho_2$, we immediately see that, even if the densities are smooth on their supports, $T$ has to be discontinuous. It was noticed by Caffarelli [38], that the right assumption to be made on the support of $\rho_2$ is convexity. Thus, regularity results regarding the convex map $\xi$ and the optimal map $T = \nabla \xi$ have been obtained from the regularity of the densities, most notably by Caffarelli [38], [37], using the Monge–Ampère equation (3.24).

**Theorem 3.24** Let $\Omega_i, \ i = 1, 2$ be two bounded, open subsets of $\mathbb{R}^d$, and let $\mu_i, \ i = 1, 2$ be two probability measures respectively on $\Omega_i$, with densities $\rho_i, \ i = 1, 2$. If those densities are bounded away from zero and infinity on their support, and if $\Omega_2$ is convex, then $\xi$ is strictly convex and $C^{1,\alpha}$ on $\Omega_1$ and hence $T \in C^\alpha$ for some universal $\alpha$. If, in addition, they are continuous, then $T \in W^{2,p}$ for every $1 \leq p < +\infty$. Moreover, if $\rho_i, \ i = 1, 2$ are $C^k$ with $k \geq 1$, then $\xi$ is $C^{k+2}$. Furthermore, if both $\Omega_i, \ i = 1, 2$ are strictly convex with smooth boundaries, the regularity of $\xi$ holds even on the boundary of $\Omega_1$. In that case, $\nabla \xi$ and $\nabla \xi^*$ are diffeomorphisms, and inverse of each other.
3.3.3 Brenier’s polar factorization and rearrangements

The Monge–Kantorovich mass transportation problem has many important applications, for instance; the Brenier’s map $T = \nabla \xi$ generalizes in some sense the notion of monotone rearrangement to the multidimensional setting. This well known fact has been investigated first by Brenier in [34], and is often known as Brenier’s polar decomposition theorem of a gradient of a convex function transport (see Chapter 3 of [115]), that reads as the following

**Theorem 3.25** Let $\Omega_i, i = 1, 2$ be a measurable subset of $\mathbb{R}^d$, and let $\mu_i \in \mathcal{P}(\Omega_i)$; respectively, with finite second–order moments $\int_{\Omega_2} |x_2|^2 \, d\mu_2(x_2) < +\infty$. Let $f : \Omega_1 \to \Omega_3 \subset \mathbb{R}^d$ be an $L^2(\mu_1)$ vector–valued mapping, and let $\mu_3 = f_\# \mu_1$. Assume that $\mu_i, i = 2, 3$ give no mass to small sets. Then there exists a unique rearrangement map $T = \nabla \xi$ of $f$ in the class of $L^2$ gradients of convex functions $\xi : \Omega_2 \to \Omega_3$ that transports $\mu_2$ to $\mu_3$, and a unique map $g : \Omega_1 \to \Omega_2$ pushes $\mu_1$ forwards to $\mu_2$, such that $f = T \circ g$. Moreover, $g$ is the unique $L^2$ projection of $f$ onto $G(\Omega_1, \Omega_2)$; the set of (all) maps $\tilde{g} : \Omega_1 \to \Omega_2$ such that $\tilde{g}_\# \mu_1 = \mu_2$.

Note that, since all elements in $G(\Omega_1, \Omega_2)$ have the same $L^2$–norm, the last statement of the theorem is equivalent to the statement that $g$ should maximize $\langle f, \tilde{g} \rangle_{L^2(\mu_1)}$ among all $\tilde{g}$’s in $G(\Omega_1, \Omega_2)$; that is, it solves $\max \{ \int f(x_1) \tilde{g}(x_1) \, d\mu_1(x_1) : \tilde{g}_\# \mu_1 = \mu_2 \}$. Note also, when we say that $f$ is “unique”, we mean that it is uniquely determined $\mu_1$–a.e. on $\Omega_1$. Furthermore, it has been known since Fréchet [67] that, any Borel map $f$ defined on $\Omega \subset \mathbb{R}^d$ can be monotonically rearranged over $\Omega$, so by “rearrangement of maps” we mean the following

**Definition 3.26** Let $f : \Omega_1 \to \Omega_2$ be a Borel measurable map between two measure spaces $(\Omega_i, \mu_i), i = 1, 2$. Another map $\tilde{f} : \Omega_1 \to \Omega_2$ is said to be a rearrangement of $f$, if the following property holds

$$
\mu_1(f^{-1}(B)) = \mu_1(\tilde{f}^{-1}(B)) \quad \text{for every Borel set } B \subset \Omega_2.
$$

(3.27)

In other words, there exists a nondecreasing function $\tilde{f} : \Omega_1 \to \Omega_2$ such that the measure $\mu_1$ of the preimages of any Borel set through $f$ and $\tilde{f}$ coincide.
Remark 3.27 Due to Definition 3.2, the definition above is equivalent to saying that whenever \( \zeta : \Omega_2 \to \mathbb{R} \) is a measurable map such that \( \zeta \circ f \in L^1(\mu_1) \), then \( \zeta \circ \tilde{f} \in L^1(\mu_1) \) and
\[
\int_{\Omega_1} \zeta(f(x_1)) \, d\mu_1(x_1) = \int_{\Omega_1} \zeta(\tilde{f}(x_1)) \, d\mu_1(x_1). \tag{3.28}
\]
If \( \mu_1(\Omega_1) < +\infty \), then (3.27) is equivalent to require that both integrals in (3.28) coincide also for any bounded measurable function \( \zeta \).

Notice that, Definition 3.26 means that "one cannot tell the difference between \( f \) and \( \tilde{f} \) by looking only at their values". For instance, \( f \) and \( \tilde{f} \) should have the same maximum and minimum, but it does not matter at which point. If \( \Omega_2 = \mathbb{R}^d \) and \( \zeta(x) = |x|^p \), we see from the definition that \( \|f\|_{L^p} = \|\tilde{f}\|_{L^p} \), for any \( p \), so Lebesgue norms are invariant under rearrangement. On the other hand, the great majority of rearrangements of a given smooth function are completely unsmooth; in particular the Sobolev norm \( \|
abla f\|_{L^p} \) are not invariant under rearrangement. The simplest way to construct rearrangements is via measure–preserving maps.

Definition 3.28 Let \( \Omega_1 \) be a given measurable space, equipped with a measure \( \mu_1 \). A measurable function \( g : \Omega_1 \to \Omega_1 \) is said to be measure–preserving if \( g_# \mu_1 = \mu_1 \); that is,
\[
\int_{\Omega_1} \zeta(x_1) \, d(g_# \mu_1)(x_1) = \int_{\Omega_1} \zeta(g(x_1)) \, d\mu_1(x_1) \quad \text{for every bounded (or } g_# \mu_1 \text{–integrable) Borel function } \zeta : \Omega_1 \to \mathbb{R}. \quad \text{In other words, for any (Borel) measurable set } B \subset \Omega_1, \text{ one has } \mu_1(g^{-1}(B)) = \mu_1(B). \quad \text{The set of all measure–preserving maps on a measure space } (\Omega_1, \mu_1) \text{ will be denoted by } G(\Omega_1).
\]

Example 3.29 Let \( \Omega \) be an open subset of \( \mathbb{R}^d \), equipped with \( \lambda := L^d|_\Omega \) (the \( d \)–dimensional Lebesgue measure \( L^d \) restricted on \( \Omega \)). By the change of variable formula (3.3), a \( C^1 \)–diffeomorphism \( g : \Omega \to \Omega \) is measure–preserving if and only if it has unit Jacobin; that is, \( |\det(\nabla g)| \equiv 1 \). Thus, the set of all diffeomorphisms have the unit Jacobin is the group of measure–preserving diffeomorphisms on \( \Omega \), and will be denoted by \( GD(\Omega) \). An important subgroup of this class, will be denoted by \( SD(\Omega) \), is the group of all diffeomorphisms \( g \) with \( \det(\nabla g) \equiv 1 \).
Note that, measure–preserving diffeomorphisms constitute but a small part of $G(\Omega)$. For instance, $GD((0,1))$ is reduced to just $\{- \text{Id}, \text{Id}\}$, but there are many elements in $G((0,1))$ which are not diffeomorphism, like $g(x) = 2x \pmod{1}$, $g(x) = 2x 1_{x \leq 1/2} + (1 - 2x) 1_{x > 1/2}$.

The following elementary proposition is a link between measure–preserving maps and rearrangements.

**Proposition 3.30** Let $(\Omega_1, \mu_1)$ be a measure space. If $g \in G(\Omega_1)$ and $\tilde{f} = f \circ g$, then $\tilde{f}$ is a rearrangement of $f$. “Conversely”, if $\tilde{f}$ is a one–to–one rearrangement of $f$, then $\tilde{f}^{-1} \circ f \in G(\Omega_1)$.

**Proof.**
1. Assume $\tilde{f} = f \circ g$, $g \in G(\Omega_1)$. Then, for every measurable $\zeta : \Omega_1 \to \mathbb{R}^+$, $\int(\zeta \circ \tilde{f}) d\mu_1 = \int(\zeta \circ f) \circ g d\mu_1 = \int(\zeta \circ f) d(g \# \mu_1) = \int(\zeta \circ f) d\mu_1$. So, $\tilde{f}$ is a rearrangement of $f$.

2. Let $\tilde{f}$ is a one–to–one rearrangement of $f$. Define $g = \tilde{f}^{-1} \circ f$ and consider any measurable nonnegative function $\zeta$ on $\Omega_1$. Then, $\int \zeta d(g \# \mu_1) = \int(\zeta \circ g) d\mu_1 = \int(\zeta \circ \tilde{f}^{-1}) \circ f d\mu_1 = \int(\zeta \circ \tilde{f}^{-1}) \circ \tilde{f} d\mu_1 = \int \zeta d\mu_1$. So, $g$ is measure–preserving. □

**Remark 3.31** In particular, when $\Omega_1 = \Omega_2 = \Omega$ is a bounded subset of $\mathbb{R}^d$ and $\Omega_3 = \mathbb{R}^d$, the same results in Theorem 3.25 above can be obtained under the sufficient assumption that, $f : \Omega \to \mathbb{R}^d$ satisfying the nondegeneracy condition:

$$|f^{-1}(N)| = 0, \text{ for any small set } N \text{ in } \mathbb{R}^d, \quad (3.29)$$

or, by an analogous assumption that, $f \# \lambda$ is absolutely continuous, where $\lambda$ is a positive rescaling of the $d$–dimensional Lebesgue measure $\mathcal{L}^d$ on $\Omega$.

As we already saw in the previous chapter, it is a mathematical way to express that fact that the volume of “element of fluid” does not change during the time evolution. Let us recapitulate the informal considerations mentioned in Section 1.1: in Lagrangian formulation, the Euler equation becomes an evolution equation for a map $t \mapsto X(t, \cdot)$, with values in the group $GD(\Omega)$ of diffeomorphisms $\Omega \to \Omega$ with unit determinant. To recall this, we shall use the letter $g$ for the trajectory map $X$. In particular, $g$ is “measure–preserving map”: it
pushes Lebesgue measure (restricted to $\Omega$) forward to itself. The physical interpretation is that the volume of a set of particles is kept constant time–evolution, which is precisely the incompressibility. Thus, we rewrite (1.3) as

$$\partial_t g(t, x) = v(t, g(t, x)) \quad \text{or} \quad v = \dot{g} \circ g^{-1}. \quad (3.30)$$

By incompressibility assumption, the Euler equation (1.6) translates into an equation on the trajectory field $t \mapsto g(t, x)$ of $\mathbb{R}^+$ into $GD(\Omega)$,

$$\partial_t g(t, x) + v_t(x) \cdot \nabla_x g(t, x) = 0. \quad (3.31)$$

This model of transport equations has been already discussed in Subsection 1.3.3.

As a one of the related facts to Brenier’s polar factorization theorem is the monotone rearrangement on the real line, which is, indeed, an essential to the main topic of this thesis. In that case, Theorem 3.25 is, in some sense, a natural generalization of the following theorem

**Theorem 3.32** Let $f : [0, 1] \to \mathbb{R}$ be an $L^p$ map, $1 \leq p < +\infty$. Then, there exists a unique nondecreasing rearrangement $\tilde{f}$ of $f$. Moreover, there exists a measure–preserving map $g : [0, 1] \to [0, 1]$ such that $f = \tilde{f} \circ g$.

Notice that, in this context, the nondegeneracy condition (3.29) means that no level set of $f$ has a positive measure. If it is not satisfied, then there is no uniqueness of a measure–preserving map $g$ in Theorem 3.32: one can always rearrange $f$ arbitrary on a level set of positive measure. Conversely, if no level set of $f$ has positive measure, then there is uniqueness of $g$.

**Remark 3.33** As a consequence of all these considerations, the quadratic case for the cost function gives a very interesting results in dimension one as we will see in the next chapter. Suppose that $\mu_1 \in \mathcal{P}(\mathbb{R})$ is nonatomic. Then every convex function is differentiable $\mu_1$–a.e., since we know that the set of nondifferentiability points of a convex function is at most countable (this is a consequence of the fact that, if $\xi$ is convex, then the intervals $(\xi'_-(x), \xi'_+(x))$, where $\xi'_-$ and $\xi'_+$ denote the left and right derivatives, are all nonempty and
disjoint when $x$ ranges among nondifferentiability points). This implies the existence on an optimal transport map for the quadratic cost between $\mu_1$ and any measure $\mu_2 \in \mathcal{P}(\mathbb{R})$. This transport map will be the derivative of a convex function, an increasing map.
Chapter 4

Unidimensional optimal transports

In this chapter, we present some useful results of the optimal transport theory on the case \(d = 1\); this allows to give a more explicit characterization of the optimal transportation plan than the general one called Brenier’s map between two measures that given in Theorem 3.19. First of all, notice that on \(\mathbb{R}\), the gradients of convex functions coincide with monotone “non-decreasing” functions. Also, subdifferentials of convex functions (equivalently, subgradients) are maximal monotone subsets of \(\mathbb{R}^2\), i.e. complete nondecreasing graphs. By definition, a subset \(G \subset \mathbb{R}^2\) is said to be monotone if \((x_1, x_2), (\tilde{x}_1, \tilde{x}_2) \in G\) implies \([x_1 \leq \tilde{x}_1 \text{ and } x_2 \leq \tilde{x}_2]\) or \([x_1 \geq \tilde{x}_1 \text{ and } x_2 \geq \tilde{x}_2]\) (note that this is the same as \((x_1 - \tilde{x}_1)(x_1 - \tilde{x}_2) \geq 0\)). Geometrically, a “complete nondecreasing graph” is nothing but the usual graph of a nondecreasing function, which possibly some vertical lines added to make this graph “continuous” (see Figure 2.4 of [115] for derivative vs. subdifferential). These lines correspond to the points \(x\) where the left and right derivatives \(\xi'_-(x)\) and \(\xi'_+(x)\) of convex function \(\xi\) do not agree. Indeed, for any \(x \in \mathbb{R}\), \(\partial \xi(x) = [\xi'_-(x), \xi'_+(x)]\).

In Section 4.1, we focus on the theory of optimal transportation on the real line, which is easier and already has many consequences. In Subsection 4.1.1, we introduce the cumulative distribution function and its generalized inverse. Subsection 4.1.2 is devoted to the optimality of transport plans on the real line for the quadratic cost. In Subsection 4.1.3, we introduce the Monotone transport maps and plans on \(\mathbb{R}\). Subsection 4.1.4 is devoted to the optimality of monotone maps for a convex cost function. In Section 4.2, we introduce another definition to one of essential tools in the unidimensional optimal transport theory: the so-called \(L^p\)-
Wasserstein distances (that is known, in the high dimensional case, as transport costs on the space of measures) induced by the $L^p$–norm of the generalized inverses of the cumulative distribution functions. We study their properties in $p^{\text{th}}$–Wasserstein spaces defined on $\mathbb{R}$; that is, the set of all Borel probability measures on the real line with finite moments of order $p$. In Subsection 4.2.1, we will introduce the $p^{\text{th}}$–Wasserstein space and $L^p$–Wasserstein distance. Subsection 4.2.2 is devoted to the topology properties of the $p^{\text{th}}$–Wasserstein space induced by $L^p$–Wasserstein distance on the real line, which is a metric space when endowed with the $L^p$–Wasserstein distance. In subsection 4.2.3, we present an alternative formulation of weak convergence in terms of the cumulative distribution functions. We refer for an extensive overview to the books of Villani [115], [116]. For a probabilistic introduction we refer the reader to the book of Ambrosio, Gigli and Savaré [17], Filippo [104].

4.1 Monotonicity of optimal transports on the real line

We already discussed, in Remark 3.33, the quadratic cost in the one–dimensional case: as soon as the source measure $\mu_1$ has no atoms, there exists an optimal map, which is monotone nondecreasing. We want to discuss and give a full characterization of the monotone (in the following by “monotone” we mean monotone nondecreasing) transport map between two given measures $\mu_i$, $i = 1, 2$ in terms of their cumulative distribution functions and generalize to the case where $\mu_1$ may have atoms. This discussion is independent of optimal transport considerations.

4.1.1 Inverse distribution functions

At significant occasions in this work, the study is reduced to the uniform distribution by means of the inverse function $F^{-1}$ (see Proposition 4.4 below). It is therefore important to freely work with the inverse distribution functions and their analytic properties. This section thus collects material on supports and continuity, modulus of continuity and absolute continuity of inverse distribution functions. We finish with a notion of $I$-function which plays an important role in the investigation of [29] where is addressed in that framework and the study of integrals containing the derivative of $F^{-1}$. 

Let us consider now probability measures on the real line. They enjoy a nice representation property known as the cumulative distribution function (c.d.f.): Given a probability measure $\mu$ on the Borel sets of $\mathbb{R}$, we define its c.d.f. $F$ associated to $\mu$ through

$$F(x) := \int_{-\infty}^{x} \mu(z) \, dz = \mu([-\infty, x]), \quad x \in \mathbb{R}. \quad (4.1)$$

Clearly, the measure $\mu$ is uniquely determined by its c.d.f. $F$, so $\mu$ (or, $F$) is said to be degenerate if $\mu$ is a Dirac mass. We sometimes denote by $X$ a random variable on a probability space $(\Omega, \Sigma, \mathbb{P})$ with distribution (law) $\mu$.

The c.d.f. $F$ is easily seen to be nondecreasing and right–continuous since if $x_k \to x^+$ as $k \to +\infty$, then $(-\infty, x] = \cap_{k \in \mathbb{N}} (-\infty, x_k]$ and hence $\mu([-\infty, x]) = \lim_{k \to +\infty} \mu([-\infty, x_k]) = \inf_{k \in \mathbb{N}} \mu([-\infty, x_k])$. Furthermore, $F(-\infty) = 0$, $F(+\infty) = 1$. It is continuous at any point where $\mu$ has no atom since if $x_k \to x^-$ as $k \to +\infty$, then $\lim_{k \to +\infty} \mu([-\infty, x_k]) = \mu([-\infty, x])$).

The opposite holds true as well: given any $\tilde{F} : \mathbb{R} \to [0, 1]$, right–continuous, nondecreasing and such that $\tilde{F}(-\infty) = 0$, $\tilde{F}(+\infty) = 1$, there exists a unique probability measure $\mu \in \mathcal{P}(\mathbb{R})$ such that $F = \tilde{F}$. Indeed, $\mu([-\infty, x]) := \tilde{F}(x)$ defines a unique Borel probability measure (because half–lines generates the whole Borel $\sigma$–algebra, the sets of the form $]-\infty, x]$ are enough to generate all open sets: we have $\mu([a, b]) = F(b) - F(a)$, $\mu([a, b]) = \sup\{F(t) - F(a) : t < b\}$ and the measure of any open set is obtained via countable disjoint unions of open intervals). Thus, the c.d.f. $F$ is important because it characterizes the measure $\mu$.

Unfortunately, the c.d.f. above cannot always be inverted, as it is not always strictly increasing, but we can define the generalized inverse of $F$ which is also known as the pseudo–inverse or inverse distribution function.

**Definition 4.1** Given a nondecreasing and right–continuous function $F : \mathbb{R} \to [0, 1]$, its generalized inverse is the (possibly discontinuous) function $F^{-1} : (0, 1) \to \mathbb{R} \cup \{\pm \infty\}$ given by

$$F^{-1}(y) := \inf \{x \in \mathbb{R} : y \leq F(x)\}, \quad (4.2)$$

where the infimum is a minimum as soon as the set is nonempty (otherwise it is $+\infty$) and
bounded from below (otherwise it is $-\infty$), thanks to right–continuity of $F$.

It is convenient furthermore to extend this function to $[0, 1]$ by monotonicity, setting

$$F^{-1}(0) = F^{-1}(0^+) = \inf \{ x \in \mathbb{R} : F(x) > 0 \}, \quad F^{-1}(1) = F^{-1}(1^-) = \sup \{ x \in \mathbb{R} : F(x) < 1 \}$$

similarly to the standard convention $F(-\infty) = 0, F(+\infty) = 1$. Thus, we have $F^{-1}(0)$ is the infimum of the support of $\mu$ and $F^{-1}(1)$ its supremum.

The next remark lists a number of basic properties and relations between $F$ and $F^{-1}$. The various claims are elementary and verified in a straightforward manner.

**Remark 4.2** A generalized inverse $F^{-1}$ is strictly increasing on $(0, 1)$ if and only if a c.d.f. $F$ is continuous. Furthermore, for every $0 < y < y' < 1$ and $x \in \mathbb{R}$, it enjoys the following (monotonicity) properties

(i) $F^{-1}(y) \leq x < F^{-1}(y')$ if and only if $y \leq F(x) < y'$.

(ii) $y \leq F \circ F^{-1}(y)$ with equality if and only if $F(z) = y$ for some $z \in \mathbb{R}$. In particular, $y = F \circ F^{-1}(y)$ if $F$ is continuous.

(iii) $x \geq F^{-1} \circ F(x)$ if and only if $F^{-1}(0) \leq x$. Moreover, $x > F^{-1} \circ F(x)$ if and only if $F(x) = F(z)$ for some $x > z$ ($F^{-1}(0) < x$).

Note that, whenever $F$ is strictly monotone and continuous, there is no problem $F^{-1}$ is strictly monotone and continuous as well, and equality holds in both expressions. If $F \equiv c$ on a given interval $J$, then $F^{-1}$ has a jump in $c$ between $\inf J = \lim_{y' \to c^-} F^{-1}(y')$ and $\sup J = F^{-1}(c)$. In this case, $F(F^{-1}(c)) = F(\sup J) = c$, but for all $x' \in J$ we have $F^{-1}(F(x')) = F^{-1}(c) = \sup J \geq x'$. On the other hand, if $F$ has a jump in $\hat{x}$ between the values $c_1 < c_2$ (and $F(x) = c_2$ by right continuity), then $F^{-1}(t c_1 + (1-t)c_2) = \hat{x}$ for all $t \in [0, 1]$, $F^{-1}$ is constant in $[c_1, c_2]$. In this case, $F^{-1}(F(\hat{x})) = F^{-1}(c_2) = \hat{x}$, but for all $\tilde{y} \in [c_1, c_2]$ we have $F(F^{-1}(\tilde{y})) = F(x) = c_2 \geq \tilde{y}$ (refer to Figure 2.5 [115]).

A simple consequence of the remark above is the following lemma.

**Lemma 4.3** If $F$ is a c.d.f., then $F^{-1}(y) \leq x$ if and only if $F(x) \geq y$ and $F^{-1}(y) < x$ if and only if $F(x) < y$. 

Proof. Since the infimum in the definition of $F^{-1}$ is attained, $F(F^{-1}(y)) \geq y$ for any $y \in [0,1]$ (thanks to Remark 4.2). Thus, if $F^{-1}(y) \leq x$ for some $x \in \mathbb{R}$, then $F(x) \geq F(F^{-1}(y)) \geq y$, as $F$ is nondecreasing. Conversely, if $F(x) \geq y$, then the definition of $F^{-1}$ implies $F^{-1}(y) \leq x$. The second assertion follows the same argument.

Note that, the both statements in the above lemma also hold for $y = 1$. On the basis of this lemma, we address the proofs of the following propositions.

The use of the generalized inverse functions is mainly explained by the following well–known observation.

**Proposition 4.4** Let $F$ be a c.d.f.. If $U$ is a random variable uniformly distributed in $(0,1)$, then the random variable $F^{-1}(U)$ has $F$ as its c.d.f..

**Proof.** The proof is immediate, since by the lemma above, for all $x \in \mathbb{R},$

$$\chi \{ y \in (0,1) : F^{-1}(y) \leq x \} = \chi \{ y \in (0,1) : F(x) \geq y \} = F(x),$$

where $\chi$ denotes Lebesgue measure $\mathcal{L}^1$ restricted on $(0,1)$. \hfill $\Box$

Another related (although less universal) property is that, when a random variable $X$ has a continuous c.d.f. $F$, the random variable $U = F(X)$ is uniformly distributed in the interval $(0,1)$. The next statement describes how the transform $F \mapsto F^{-1}$ acts between its domain and image.

**Proposition 4.5** Any generalized inverse function is nondecreasing and right–continuous. Moreover, for any nondecreasing, right–continuous function $G : (0,1) \to \mathbb{R}$, there exists a unique c.d.f. $F$ such that $F^{-1} = G$.

**Proof.** Assume first that $F^{-1}$ is not right–continuous at some point $y$. That is, setting $x = F^{-1}(y)$, there exists a $\delta > 0$ such that $F^{-1}(y + \varepsilon) \geq F^{-1}(y) + \delta = x + \delta$ for all $\varepsilon > 0$. By Lemma 4.3, this is equivalent to saying that $F(x + \delta) \leq y + \varepsilon$. Letting $\varepsilon \to 0$, we get $F(x + \delta) \leq y$, which in turn is equivalent to $x + \delta \leq F^{-1}(y)$, a contradiction. The first assertion of the proposition is established.

Next turn to the existence part of the second claim. Set $G(1) = G(1^-)$ and define on $\mathbb{R}$ the function $F(x) = \sup \{ y \in [0,1] : x \geq G(y) \}$, using (if necessary) the convention that $\sup \emptyset = 0$. Then $F(x) = x$ for all $x \in (0,1)$, hence $F^{-1}(x) = x$ for all $x \in (0,1)$. For $x \in (0,1)$, we have $F^{-1}(x) = x$, as desired. \hfill $\Box$
0. In particular, \( F(x) > 0 \) if and only if \( x \geq G(y) \) for some \( y \in (0, 1) \). By construction, \( F \) is nondecreasing and takes values in \([0, 1]\). To prove that it is right–continuous, observe the following. Fix \( y_0 \in (0, 1) \) and \( x \in \mathbb{R} \). If \( x \geq G(y_0) \), then the set \( \{ x \geq G(y) \} \) is nonempty, and \( y_0 \geq F(x) \). Conversely, if \( y_0 \geq F(x) \), then \( F(x) > 0 \), so the set \( \{ x \geq G(y) \} \) is nonempty. Hence, using the right continuity of \( G \), \( \sup \{ y \in (0, 1) : x \geq G(y) \} \leq y_0 \). Then, there exists \( y_0 \geq y \) and \( x \geq G(y) \) which implies \( x \geq G(y_0) \). Thus, for all \( y \in (0, 1) \) and \( x \in \mathbb{R} \),

\[
x \geq G(y) \implies F(x) \geq y.
\]

Now, assume that \( F \) is not right–continuous at some point \( x_0 \) and put \( y_0 = F(x_0) \). Then, there exists \( \delta > 0 \) such that \( F(x_0 + \varepsilon) \geq F(x_0) + \delta = y_0 + \delta \) for all \( \varepsilon > 0 \). In particular, \( y = y_0 + \delta \in (0, 1] \). By (4.3), \( G(y_0 + \delta) \leq x_0 + \varepsilon \). Letting \( \varepsilon \to 0 \), we are led to \( G(y_0 + \delta) \leq x_0 \), which in turn is equivalent to \( y_0 + \delta \leq F(x_0) \), a contradiction with \( y_0 = F(x_0) \). The existence of the distribution function \( F \) is therefore established.

We are left with uniqueness. Let \( \tilde{F} \) be another distribution function such that \( \tilde{F}^{-1} = G \). By (4.3) applied to both \( F \) and \( \tilde{F} \), we have \( \tilde{F}(x) \geq y \) if and only if \( F(x) \geq y \) for all \( y \in (0, 1] \) and \( x \in \mathbb{R} \), which clearly amounts to \( \tilde{F}(x) = F(x) \). The proof is therefore complete.

Notice that, when thinking of \( F \) in terms of \( F^{-1} \), there should be no constraint on the latter function except for the property of being nondecreasing and right–continuous. The inverse function \( F^{-1} \) of a given c.d.f. \( F \) generates a nonnegative Borel measure \( \mu^{-1} \) on \((0, 1)\), defined for semi–open intervals by \( \mu^{-1}([y, y')) = F^{-1}(y') - F^{-1}(y) \) for \( 0 < y < y' < 1 \). It may be called the inverse measure (with respect to the probability measure \( \mu \) with the c.d.f. \( F \)). The next statement describes \( \mu^{-1} \) for any Borel sets.

**Proposition 4.6** Any nondegenerate c.d.f. \( F \) restricted to the interval \( J = \{ x \in \mathbb{R} : 0 < F(x) < 1 \} \) pushes forward the Lebesgue measure \( \mathcal{L}^1 \) on \( J \) onto the inverse measure \( \mu^{-1} \). That is, for any Borel set \( B \subset (0, 1) \), \( \mathcal{L}^1 \{ x \in J : F(x) \in B \} = \mu^{-1}(B) \).

**Proof.** It is sufficient to verify this equality for \( B = [y, y') \) with arbitrary \( 0 < y < y' < 1 \) and indeed, by Remark 4.2,

\[
\lambda \{ x \in J : y \leq F(x) < y' \} = \mu^{-1}([y, y']) = \mu^{-1}(\{ x \in J : F^{-1}(y) \leq x < F^{-1}(y') \}).
\]
where \( \lambda \) denotes Lebesgue measure \( \mathcal{L}^1 \) restricted on \( J \).

We comment now on issues connected with the support of measures. Given a c.d.f. \( F \) on \( \mathbb{R} \) and its inverse function \( F^{-1} \), observe first that \( [F^{-1}(0), F^{-1}(1)] \) represents the smallest closed interval in \([ -\infty, +\infty ]\) on which the measure \( \mu \) with the distribution function \( F \) is supported. In general, the support \( \text{spt}(\mu) \) is defined as the smallest closed subset of the real line \( \mathbb{R} \) of full \( \mu \)-measure. This set can be defined as the collection of all points \( x \) of growth of \( F \), i.e. such that \( F(x + \varphi) > F(x - \varepsilon) \) for every \( \varepsilon > 0 \). A similar definition is applied to a general Borel measure generated by a nondecreasing function on a given interval, and in particular to the inverse measure \( \mu^{-1} \) on \((0, 1)\) generated by the inverse distribution function \( F^{-1} \). Its support, i.e. the smallest closed subset of \((0, 1)\) of full \( \mu^{-1} \)-measure, is described in terms of the image set \( \text{Im}(F) = \{ F(x) : -\infty \leq x \leq +\infty \} \). As an equivalent definition, one may involve the inverse function to write the representation

\[
\text{Im}(F) \cap (0, 1) = \left\{ y \in (0, 1) : F^{-1}(y) < F^{-1}(y') \text{ for all } y' \in (y, 1) \right\}.
\]  

(4.4)

Indeed, using Remark 4.2, we have \( F^{-1}(y) < F^{-1}(y') \) if and only if there exists \( x \in \mathbb{R} \) such that \( F^{-1}(y) \leq x < F^{-1}(y') \), i.e. \( \exists x \in \mathbb{R}, y \leq F(x) < y' \). By the continuity of \( F \) from the right, the latter property holds true for all \( y' \in (y, 1) \) if and only if \( y = F(x) \) for some \( x \in \mathbb{R} \), thus proving the claim.

By Proposition 4.6, since \( F : \mathbb{R} \to \text{Im}(F) \), the measure \( \mu^{-1} \) is supported on \( \text{Im}(F) \) (once we realize that the image set is Borel measurable). This set does not need be closed, but its closure is just \( \text{clos}(\text{Im}(F)) = \text{Im}(F) \cup \{ F(x^-) : x \in \mathbb{R} \} \).

**Proposition 4.7 (Support of the inverse measure)** Let \( \mu \) be a probability measure on \( \mathbb{R} \) with c.d.f. \( F \). A number \( y \in (0, 1) \) is a point of growth of \( F^{-1} \) if and only if \( y = F(x) \) or \( y = F(x^-) \) for some \( x \in \mathbb{R} \). Equivalently, \( \text{spt}(\mu^{-1}) = \text{clos}(\text{Im}(F)) \cup (0, 1) \).

**Proof.** The support of \( \mu^{-1} \), \( \text{spt}(\mu^{-1}) \), represents the collection of all points \( t \) of growth of \( F^{-1} \) on \((0, 1)\), i.e. such that \( F^{-1}(y_0) < F^{-1}(y_1) \) whenever \( 0 < y_0 < y < y_1 < 1 \). Using Remark 4.2 and arguing as before, we have \( F^{-1}(y_0) < F^{-1}(y_1) \) if and only if there exists \( x \in \mathbb{R} \) such that \( F^{-1}(y_0) \leq x < F^{-1}(y_1) \), i.e. \( \exists x \in \mathbb{R}, y_0 \leq F(x) < y_1 \), or
Let \([y_0, y_1] \cap \text{Im}(F) \neq \emptyset\). Hence, \(y\) is a point of growth of \(F^{-1}\), if and only if \([y_0, y_1] \cap \text{Im}(F) \neq \emptyset\) for all \(y_0, y_1\) such that \(y_0 < y < y_1\). But the latter is equivalent to the property \(y \in \text{clos}(\text{Im}(F))\).

Next, we comment more on the structure of the image set \(\text{Im}(F)\). If the measure \(\mu\) with c.d.f. \(F\) is nonatomic (\(F\) is continuous), then \(\text{Im}(F) = [0, 1]\), and \(F^{-1}\) is strictly increasing.

In the general case, let us return to the representation (4.4) and consider the complement of the image in \((0, 1)\), \(I = (0, 1) \setminus \text{Im}(F) = \{y \in (0, 1) : F^{-1}(y) = F^{-1}(y') \text{ for some } y' \in (y, 1)\}\).

With every point \(y\) in \(I\), this set also contains some nonempty interval \([y, y')\). For a rational number \(r \in (0, 1)\), denote by \(I_r\) the union of all such intervals that contain \(r\). Clearly, if \(I_r\) is nonempty, it is an interval either of type \([a, b)\) or \((a, b]\). Thus, \(I = \bigcup_r I_r\), which shows in particular that \(\text{Im}(F)\) is always Borel measurable. The following proposition collects conditions insuring the continuity of the generalized inverse functions in terms of the support of the measure.

**Proposition 4.8 (Continuity and support)** Let \(\mu\) be a probability measure on \(\mathbb{R}\) with c.d.f. \(F\). The following properties are equivalent:

(i) The c.d.f. \(F\) is strictly increasing on the interval \(J_0 = \{x \in \mathbb{R} : 0 < F(x) < 1\}\);

(ii) The inverse function \(F^{-1}\) is continuous;

(iii) The inverse measure \(\mu^{-1}\) is nonatomic;

(iv) The support of \(\mu\) is a closed interval on the real line, finite or not.

Thus, for the continuity of \(F^{-1}\), the support \(J = \text{spt}(\mu)\) should be one of the following types: \((−\infty, +\infty)\), \((−\infty, b]\), \([a, +\infty)\), or \([a, b]\) with some finite \(a, b\). Then, in the corresponding of \(J_0\) cases: \((−\infty, +\infty)\), \((−\infty, b]\), \([a, +\infty)\) or \((a, +\infty)\), and \([a, b]\) or \((a, b]\) (in the two last cases depending on whether \(\mu\) has an atom at the point \(a)\). Anyhow, all points of \(J\) will be points of growth of \(F\). In particular, according to Remark 4.2, \(F^{-1} \circ F(x) = x\) for every \(x \in J\).

**Proof of Proposition 4.8.** The equivalence of (ii) and (iii) is standard. For (ii) \(\Rightarrow\) (i), assume by contradiction that \(F\) is not strictly increasing on \(J_0\), that is, \(y = F(x') = F(x)\)
for some \( x' < x \) with \( 0 < y < 1 \). Let \( x' \) be the smallest number satisfying this equality with fixed \( x \). Then, by Remark 4.2, \( F^{-1}(y) = x' \), while \( F^{-1}(y') > x \) whenever \( y' > y \). Hence, \( F^{-1} \)

is discontinuous at \( y \), thus proving the implication. Conversely, if \( F^{-1} \) is discontinuous at \( y \in (0, 1) \) and \( x' = F^{-1}(y) \), then \( x = F^{-1}(y^+) > x' \). Hence \( F \) is constant on \([x', x)\) proving the implication \( (i) \Rightarrow (ii) \). Finally, assuming \( (i) \), any point in \( J_0 \) is a point of growth of \( F^{-1} \). Hence, \( \text{spt}(\mu) = \text{clo}(J_0) \) which implies \( (iv) \). In turn, \( (iv) \) implies that \( \mu(x - \varepsilon, x + \varepsilon) > 0 \) for every \( x \in J \) and \( \varepsilon > 0 \), and we arrive at \( (i) \). The proof is complete. \( \square \)

To conclude this paragraph, we illustrate the preceding results with the following example.

**Example 4.9** For the mass point \( \mu = \delta_x \), \( x \in \mathbb{R} \) (the degenerate case), \( F^{-1}(y) = x \) for \( y \in (0, 1) \). Hence \( \mu^{-1} = 0 \). For the “Bernoulli” measure \( \mu = p\delta_{x_1} + (1 - t)\delta_{x_2} \), for \( x_1 < x_2 \) and \( t \in (0, 1) \), \( F^{-1}(y) = x_1 \) if \( y \in (0, t] \) and \( F^{-1}(y) = x_2 \) if \( y \in (t, 1) \). Hence \( \mu^{-1} = (x_2 - x_1)\delta_t \) which is a multiple of the mass point. Let \( \chi \) denote the uniform measure on \((0, 1)\). For a mixture of the Bernoulli and the uniform measure \( \mu = (1/4)\delta_0 + (1/2)\chi + (1/4)\delta_1 \), the inverse distribution function is continuous and is given by \( F^{-1}(y) = 0 \) if \( y \in (0, 1/4] \), \( F^{-1}(y) = 2y - (1/2) \) if \( y \in (1/4, 3/4] \) and \( F^{-1}(y) = 1 \) if \( y \in (3/4, 1) \). In this case \( \mu^{-1} \) represents a multiple of the uniform distribution on \([1/4, 3/4]\).

Once the inverse function \( F^{-1} \) of a c.d.f. \( F \) is continuous, one can try to quantify this property by considering its *modulus of continuity* \( \omega_{F^{-1}}(\varepsilon) = \sup\{|F^{-1}(y) - F^{-1}(y')|: |y - y'| \leq \varepsilon, \, y, y' \in (0, 1)\} \) for \( \varepsilon \in (0, 1] \), which is an optimal function \( \omega \) such that \( |F^{-1}(y) - F^{-1}(y')| \leq \omega(|y - y'|) \) for every \( y, y' \in (0, 1) \). However, as is made clear by the next statement, the study of moduli of continuity is restricted to the class of compactly supported measures.

**Proposition 4.10** Let \( \mu \) be a probability measure on \( \mathbb{R} \) with c.d.f. \( F \). Then, \( \omega_{F^{-1}}(\varepsilon) < +\infty \) for some (or, every) \( \varepsilon \in (0, 1) \) if and only if \( \mu \) is compactly supported. Moreover, if \( \mu \) is supported on an interval of length \( l \), then \( \omega_{F^{-1}} < l \). Furthermore, \( F^{-1} \) is continuous if and only if \( \omega_{F^{-1}}(0+) = 0 \).

The statement is obvious. Let us only stress that \( \mu \) is not compactly supported if and only if \( F^{-1}(0) = -\infty \) or \( F^{-1}(1) = +\infty \). In the latter case, we have \( \omega_{F^{-1}}(\varepsilon) \geq F^{-1}(1) - F^{-1}(1 - \varepsilon) = +\infty \) for every \( \varepsilon \in (0, 1) \).
If \( \mu \) is compactly supported, and \([a_1, a_2]\) is the smallest segment where \( \mu \) is supported, the behaviour of the modulus of continuity \( \omega_{F^{-1}} \) near zero can be connected with the dual notion— an analogous “modulus of increase” of the c.d.f \( F \). This function may be defined as \( \varepsilon_F(\delta) = \inf \{ F(x_2) - F(x_1) : x_2 - x_1 > \delta, \; x_1, x_2 \in [a_1, a_2] \} = \inf \{ F(x_2) - F(x_1-) : x_2 - x_1 \geq \delta, \; x_1, x_2 \in [a_1, a_2] \} \) for every \( \delta \in (0, a_2 - a_1) \).

**Proposition 4.11** Let \( F \) be the c.d.f associated with a probability measure \( \mu \) such that \( \text{spt}(\mu) = [a_1, a_2] \). Then, for every \( \varepsilon \in (0, 1) \) and \( \delta \in (0, a_2 - a_1) \), \( \omega_{F^{-1}}(\varepsilon) \leq \delta \) if and only if \( \varepsilon_F(\delta) \geq \varepsilon \). In particular, \( \varepsilon_F(\delta) = \inf \{ 0 < \delta < a_2 - a_1 : \varepsilon_F(\delta) \geq \varepsilon \} \).

**Proof.** The support assumption means that the inverse function \( F^{-1} \) is continuous on \((0, 1)\) according to Proposition 4.8. We may assume that \( \mu \) is non-degenerate so that \( a_1 < a_2 \). By definition, \( \varepsilon_F(\delta) \geq \varepsilon \) means that \( s - t \leq \varepsilon \) implies \( F^{-1}(y_2) - F^{-1}(y_1) \leq \delta \) whenever \( y_1 < y_2 \) in \((0, 1)\). Since \( F^{-1} \) is continuous, this implication may be rewritten as \( y_2 - y_1 < \varepsilon \) implies \( F^{-1}(y_2) - F^{-1}(y_1) \leq \delta \). Moreover, this description will not change if we require additionally that \( F^{-1}(y_1) < F^{-1}(y') < F^{-1}(y_2) \) for every \( y_1 < y' < y_2 \). Indeed, otherwise, the interval \((y_1, y_2)\) may be decreased without change of the value \( F^{-1}(y_2) - F^{-1}(y_1) \). As explained in the proof of Proposition 4.4, such a requirement is equivalent to the property that \([y_1, y') \cap \text{Im}(F) \neq \emptyset \) and \([y', y_2) \cap \text{Im}(F) \neq \emptyset \), for every \( y' \in (y_1, y_2) \). But then \( y_1 = F(x_1) \) and \( y_2 = F(x_2-) \) for some \( x_1 < x_2 \) in \([a_1, a_2] \), and hence \( F^{-1}(y_1) = x_1 \). In addition, by the right–continuity of \( F^{-1} \), \( F^{-1}(y_2) = F^{-1}(F(x_2-)) = \lim_{x \to x_2} F^{-1}(F(x)) = x_2 \) since \( F^{-1}(F(x)) = x \) for all \( x \in [a_1, a_2] \). Thus, the inequality \( \varepsilon_F(\delta) \geq \varepsilon \) is reduced to the statement \( F(x_2) - F(x_1) < \varepsilon \) implies \( x_2 - x_1 \leq \delta \) (for \( x_1 < x_2 \) in \([a_1, a_2] \)), or equivalently, \( x_2 - x_1 > \delta \) implies \( F(x_2) - F(x_1) \geq \varepsilon \). The latter amounts to \( \varepsilon_F(\delta) \geq \varepsilon \). \( \square \)

As in the preceding paragraph, we conclude with the following example illustrating these results.

**Example 4.12** Let a random variable \( X \) have a unimodal distribution, symmetric about the point 1/2, with support \([0, 1]\). The latter means that the distribution function \( F \) of \( X \) is convex on \([0, 1/2]\) and concave on \([1/2, 1]\), with the symmetry property \( F(1 - x) = 1 - F(x) \) for all \( x \in [0, 1] \) (for simplicity, let us exclude the case where \( F \) has a jump at the point
Then, the probabilities \( P\{x_2 \leq X \leq x_1\} = F(x_2) - F(x_1) \), for \( x_2 - x_1 \geq \delta \) and \( x_1, x_2 \in [0, 1] \), are minimized for \( x_2 = 0, x_1 = \delta \), so \( \varepsilon F(\delta) = F(\delta) \). Hence, \( \omega_{F^{-1}}(\varepsilon) = F^{-1}(\varepsilon) \) for every \( \varepsilon \in (0, 1) \). This function is concave on the interval \([0, 1/2]\) and convex on \([1/2, 1]\).

Now we consider absolutely continuous property for the inverse function \( F^{-1} \) on \((0, 1)\) (see Appendix A.5.4). First, let us state one immediate important consequence of the absolute continuity assumption on \( F^{-1} \).

**Proposition 4.13** If \( F \) is a nondegenerate c.d.f. such that \( F^{-1} \) is absolutely continuous, then the image set \( \text{Im}(F) \) has a positive Lebesgue measure.

**Proof.** The nondegeneracy of \( F \) insures that \( F^{-1} \) generates a nonzero inverse measure \( \mu^{-1} \) where \( \mu \) is the probability measure associated with \( F \). By the second assumption, the inverse measure is absolutely continuous with respect to the Lebesgue measure \( \chi \). So, \( \chi(\text{Im}(F)) = 0 \) would imply that \( \mu^{-1}(\text{Im}(F)) = 0 \) which is impossible since \( \mu^{-1} \) is supported on \( \text{Im}(F) \). □

Next, we turn to a full characterization of the absolute continuity of \( F^{-1} \) in terms of the distribution function \( F \). In general, the measure \( \mu \) generated by \( F \) admits a unique decomposition \( \mu = \mu_0 + \mu_1 + \mu_2 \), where \( \mu_0 \) is a discrete measure, \( \mu_1 \) is a singular continuous measure which is orthogonal to the Lebesgue measure \( \lambda := L^1 \) on \( \mathbb{R} \), and \( \mu_2 \) is a measure which is absolutely continuous with respect to \( \lambda \). They are respectively called the discrete component, the singular continuous component, and the absolutely continuous component of \( \mu \). Furthermore, by the Lebesgue differentiation theorem, the limit \( f(x) = \lim_{x' \to x}(F(x') - F(x))/(x' - x) \), for \( x' \neq x \), exists and is finite for almost all \( x_1 \), and represents the density of \( \mu_2 \) with respect to Lebesgue measure. That is, \( f(x) \, dx = d\mu_2(x) \) in the sense of Measure Theory (Radon-Nikodým derivative).

**Proposition 4.14 (Characterization of absolute continuity of \( F^{-1} \))** Let \( \mu \) be a nondegenerate probability measure on \( \mathbb{R} \) with c.d.f. \( F \). The inverse function \( F^{-1} \) is absolutely continuous on \((0, 1)\) if and only if \( \mu \) is supported on an interval, finite or not, and the absolutely continuous component of \( \mu \) has on that interval an a.e. positive density (with respect to the Lebesgue measure).
Since the absolute continuity is stronger than just continuity, necessarily the support of \( \mu \) should be a closed interval \( J = [-\infty, +\infty] \), as already indicated in Proposition 4.8. In that case, an additional requirement concerning the density which is needed for the absolute continuity of \( F^{-1} \) is equivalent to the property that the Lebesgue measure on \( J \) is absolutely continuous with respect to \( \mu \).

Proof of Proposition 4.14. We may assume that \( F^{-1} \) is continuous on \((0,1)\), so that \( J = \text{spt}(\mu) \) is an interval (not shrinking to a point by the nondegeneracy assumption). In particular, \( F^{-1}(F(x)) = x \), for every \( x \in J \). By definition, \( F^{-1} \) is absolutely continuous on \((0,1)\) if and only if, for every \( 0 < a_1 < a_2 < 1 \) and \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that, for any sequence of nonoverlapping intervals \((a_1^k, a_2^k) \subset [a_1, a_2] \), \( \sum_k (a_2^k - a_1^k) < \delta \) implies \( \sum_k (F^{-1}(a_2^k) - F^{-1}(a_1^k)) < \varepsilon \). Note that when a continuous function \( f \) is nondecreasing, in the definition of the absolute continuity (see Appendix A.5.4) one may require without loss of generality that \( f(a_1^k) < f(t) < f(a_2^k) \) for \( a_1^k < t < a_2^k \) (otherwise, the intervals \((a_1^k, a_2^k)\) may be decreased without change of the value \( f(a_2^k) - f(a_1^k) \)). In the case \( f = F^{-1} \), as was already explained in the proof of Proposition 4.11, such a requirement implies that \( a_1^k = F(x_1^k) \) and \( a_2^k = F(x_2^k) \) for some \( x_1^k < x_2^k \) in \( J \) and, moreover, \( F^{-1}(a_1^k) = x_1^k \) and \( F^{-1}(a_2^k) = x_2^k \). Thus, the definition of the absolute continuity of \( F^{-1} \) reduces to the statement that, for any finite interval \([x_1, x_2] \subset J \) and any \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that, for any sequence of nonoverlapping intervals \((x_1^k, x_2^k) \subset [x_1, x_2] \), \( \sum_k (F(x_2^k) - F(x_1^k)) < \delta \) implies \( \sum_k (x_2^k - x_1^k) < \varepsilon \). Equivalently, if \( (A) < \delta \) then \( \text{mes}(A) < \varepsilon \), for the open set \( A = \bigcup_k (x_1^k, x_2^k) \).

Using regularity of measures, this implication can easily be extended to the class of all Borel subsets \( A \) of \([x_1, x_2]\). Therefore, the Lebesgue measure on \([x_1, x_2]\) is absolutely continuous with respect to the measure \( \mu \) restricted to \([x_1, x_2]\). Extending \([x_1, x_2]\) to the whole support interval, we finally conclude that \( F^{-1} \) is absolutely continuous on \((0,1)\) if and only if the Lebesgue measure on \( J \) is absolutely continuous with respect to \( \mu \). \( \square \)

In case \( \mu \) is absolutely continuous, a more precise statement is available. When \( F \) has a positive continuous derivative \( f \) in a neighbourhood of \( F^{-1}(y) \), then \( F^{-1} \) is differentiable at \( y \) and has derivative \( (F^{-1})'(y) = 1/f(F^{-1}(y)) \). In a more relaxed form, the following statement is valid.
Proposition 4.15 Let $\mu$ be a probability measure on $\mathbb{R}$ supported on an open interval, finite or not, where it has an a.e. positive density $f$, and let $F$ be the c.d.f. of $\mu$. Then, the inverse function $F^{-1}$ is strictly increasing, absolutely continuous and, moreover, for all $0 < y_0 < y_1 < 1$,

$$F^{-1}(y_1) - F^{-1}(y_0) = \int_{y_0}^{y_1} \frac{1}{f(F^{-1}(y))} \, dy.$$

In particular, a.e. $F^{-1}$ is differentiable and has derivative $(F^{-1})'(y) = 1/f(F^{-1}(y))$.

Proof. Let $\mu$ be supported on $(a_1, a_2) \subset \mathbb{R}$. Since $f(x) > 0$ a.e. on this interval, $F$ is continuous and strictly increasing on $(a_1, a_2)$, and so is the inverse function $F^{-1} : (0, 1) \to (a_1, a_2)$. Now, if a random variable $U$ is uniformly distributed in $(0, 1)$, the random variable $X = F^{-1}(U)$ has the c.d.f. $F$ and the density $f$. Hence,

$$\int_{y_0}^{y_1} \frac{1}{f(F^{-1}(y))} \, dy = \mathbb{E} \left( \frac{1}{f(F^{-1}(U))} \mathbb{1}_{\{y_0 < U < y_1\}} \right) = \mathbb{E} \left( \frac{1}{f(F^{-1}(U))} \mathbb{1}_{\{F^{-1}(y_0) < F^{-1}(U) < F^{-1}(y_1)\}} \right)$$

$$= \mathbb{E} \left( \frac{1}{f(X)} \mathbb{1}_{\{F^{-1}(y_0) < X < F^{-1}(y_1)\}} \right) = \int_{y_0}^{y_1} \frac{1}{f(x)} \mathbb{1}_{\{F^{-1}(y_0) < x < F^{-1}(y_1)\}} \, f(x) \, dx$$

$$= F^{-1}(y_1) - F^{-1}(y_0),$$

and then Proposition 4.15 is established.

In the general case, when $\mu$ has a nonzero absolute continuous component, but also may have a nonzero discrete or continuous singular component, the Radon-Nikodým derivative of $F^{-1}$ can be expressed in a similar way. Anyway, once $F^{-1}$ is absolutely continuous, for this derivative we may take the function

$$(F^{-1})'(y) = \lim_{y' \to y, \; y' > y} \frac{F^{-1}(y') - F^{-1}(y)}{y' - y} \tag{4.5}$$

by the Lebesgue differentiation theorem (see Appendix A.5). If $y \not\in \operatorname{Im}(F)$, then $F^{-1}(y') = F^{-1}(y)$ for some $y' > y$, so $(F^{-1})'(y) = 0$. Otherwise, $y = F(x)$ for some $x$ from the support $J$ of $\mu$, and $F^{-1}(y) = x$. Let us see that the above lim inf may be taken along the values $y' = F(x'-\epsilon)$ with $x' > x$. Indeed, in case $y'$ is not of the type $F(x'-\epsilon)$, we would have $F^{-1}(y'') = F^{-1}(y')$ for some $y < y'' < y'$, and then $(F^{-1}(y'') - F^{-1}(y))/(y'' - y) > (F^{-1}(y') - F^{-1}(y))/(y' - y)$. Hence, one may exclude such points $y'$ from the lim inf in (4.5).
That is, it suffices to consider the values $y' = F(x' -)$ with $x' \in J$, $x' > x$. In that case, by the right-continuity of the inverse function, necessarily $(F^{-1})'(y') = x'$, so that

$$(F^{-1})'(y) = \lim_{x' \to x, x' > x} \frac{x' - x}{F(x' -) - F(x)} = \lim_{x' \to x, x' > x} \frac{x' - x}{F(x') - F(x)} = \frac{1}{\limsup_{x' \to x, x' > x} \frac{F(x') - F(x)}{x' - x}}.$$

The following proposition summarizes the conclusion at this point.

**Proposition 4.16** Let $F$ be a nondegenerate c.d.f.. If the inverse function $F^{-1}$ is absolutely continuous, then it has the Radon-Nikodým derivative $(F^{-1})'(y') = 1/f(F^{-1}(y))$ for every $y \in \text{Im}(F) \cap (0, 1)$, where $f(x) = \limsup_{x' \to x, x' > x} (F(x') - F(x))/(x' - x)$.

Here, the function $f$ represents a specific representative of the density of the absolutely continuous component of the measure $\mu$ generated by $F$. According to Proposition 4.14, the assumption that $F^{-1}$ is absolutely continuous is equivalent to saying that $f$ is a.e. positive on the supporting interval for $\mu$. For $y \notin \text{Im}(F)$, one may set $(F^{-1})'(y') = 0$. Recall that the measure $\mu^{-1}$ generated by $F^{-1}$ is supported on the set $\text{Im}(F) \cap (0, 1)$, so it does not matter how to define the Radon-Nikodým derivative on its complement. The preceding proposition emphasizes the concept of $I$-function associated to a c.d.f. $F$ on the real line, extensively used throughout [29] investigation, i.e. a function of a c.d.f $F$, whose inverse function $F^{-1}$ is absolutely continuous on $(0, 1)$, that is defined as $I_{F}(y) = 1/(F^{-1})'(y)$ for $y \in (0, 1)$. This function is well-defined a.e., and then $(F^{-1})'$ denotes the corresponding Radon-Nikodým derivative. In particular, if an absolutely continuous probability measure on $\mu$ on $\mathbb{R}$ with distribution function $F$ is supported on an open interval and has there an a.e. positive density $f$, then the $I$–function is well-defined and is given, according to Proposition 4.15, by $I_{F}(y) = f (F^{-1}(y))$ a.e.. According to Proposition 4.16, the formula $I_{F}(y) = f (F^{-1}(y))$ remains to hold in the general case of the absolutely continuous inverse function $F^{-1}$ – however for a specific representative of the density of the absolutely continuous component of $\mu$. Nevertheless, a number of important relations and integrals containing the $I$–functions may be expressed explicitly in terms of $f$ and do not depend on how we choose the density $f$. Some of them are considered in [29].
Remark 4.17 A similar representation can be given to probability measures on $\mathbb{R}^2$: Given $\gamma \in \mathcal{P}(\mathbb{R}^2)$ and set
\[
\mathcal{R}(y_1, y_2) := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \leq y_1, x_2 \leq y_2\},
\] we define the joint cumulative distribution function associated to $\gamma$ as
\[
H(y_1, y_2) := \int_{\mathcal{R}(y_1, y_2)} d\gamma(x_1, x_2) = \gamma(\mathcal{R}(y_1, y_2)).
\]

Vice versa, for any $\tilde{H} : \mathbb{R}^2 \to \mathbb{R}$ such that $\tilde{H}$ is nondecreasing and right–continuous in both its variables, and its limits at $(-\infty, -\infty), (+\infty, +\infty)$ are respectively 0 and 1, we find a unique measure $\gamma \in \mathcal{P}(\mathbb{R}^2)$ such that $H = \tilde{H}$ (because rectangles $\mathcal{R}$ generates the whole Borel $\sigma$–algebra in $\mathbb{R}^2$). In the $d$–dimensional Euclidean space $\mathbb{R}^d$ the situation is trickier, but the knowledge of $\mu((-\infty, x_1] \times \ldots \times (-\infty, x_d])$ for every $x = (x_1, \ldots, x_d)$ is enough, as in 1D, to characterize $\mu$ (by finite differences one could get all the semi-open rectangles $(x_1^-, x_1^+] \times \ldots \times (x_d^-, x_d^+]$ and by countable disjoint unions all the open sets).

4.1.2 The optimality of transport plans for the quadratic cost

For probability measures on $\mathbb{R}^2$, there holds the following characterization theorem in $\mathbb{R}$ due to Hoeffding and Fréchet (for more details on this theorem, check [99]).

Theorem 4.18 (Hoeffding–Fréchet Theorem) Let $H : \mathbb{R}^2 \to \mathbb{R}$ be the two–dimensional c.d.f. of a probability measure $\gamma$ defined by (4.7) and let $\mu_i, i = 1, 2$ be probability measures on $\mathbb{R}$ associated with c.d.f. $F_i$, respectively. Then $\gamma \in \Gamma(\mu_1, \mu_2)$ if and only if for all $(x_1, x_2) \in \mathbb{R}^2$
\[
\sum_{i=1}^{2} F_i(x_i) - 1 \leq H(x_1, x_2) \leq \min\{F_i(x_i) : i = 1, 2\}
\]

We can now state the solution to optimal transportation problem for a quadratic cost on $\mathbb{R}$ in terms of c.d.f..

Theorem 4.19 For $i = 1, 2$ let $\mu_i \in \mathcal{P}(\mathbb{R})$ and let $F_i$ be the c.d.f. associated to them, respectively. Let $\gamma \in \mathcal{P}(\mathbb{R}^2)$ be the probability measure associated to the joint c.d.f. $H(x_1, x_2) := \int \int d\gamma(x_1, x_2)$, then $\gamma \in \Gamma(\mu_1, \mu_2)$ if and only if for all $(x_1, x_2) \in \mathbb{R}^2$
\[
\sum_{i=1}^{2} F_i(x_i) - 1 \leq H(x_1, x_2) \leq \min\{F_i(x_i) : i = 1, 2\}.
\]
min\{F_i(x_i) : i = 1, 2\} on \mathbb{R}^2. Then \(\gamma \in \Gamma(\mu_1, \mu_2)\) is an optimal transportation plan for Kantorovich optimization problem (KP) between \(\mu_1\) and \(\mu_2\) with the cost \(c(x_1, x_2) = |x_1 - x_2|^2\). Moreover, the optimal transportation cost is given by

\[ T_2(\mu_1, \mu_2) := \int_0^1 |F_1^{-1}(y) - F_2^{-1}(y)|^2 \, dt. \]  

(4.8)

Proof. Notice that one could prove this theorem without using Theorem 3.19. Nevertheless, we prefer to see it as a consequence of Theorem 3.19, to show the links between the general \(\mathbb{R}^d\) case and this one-dimensional setting.

(a) First of all, due to the choice of \(H\), \(\gamma \in \Gamma(\mu_1, \mu_2)\) is a consequence of Theorem 4.18.

(b) We claim that

\[ \text{spt}(\gamma) \subset \{(x_1, x_2) \in \mathbb{R}^2 : F_1(x_1^-) \leq F_2(x_2) \text{ and } F_2(x_2^-) \leq F_1(x_1)\}, \]  

(4.9)

where \(F_i(x_i^-), i = 1, 2\) are the two left limits of \(F_i\) in \(x_i\), respectively (by monotonicity these limits always exist).

Let \((x_1, x_2)\) such that \(F_1(x_1^-) > F_2(x_2)\), being the case \(F_2(x_2^-) > F_1(x_1)\) similar (actually, the same up to a change of role between \(x_1\) and \(x_2\)). Fixed \(\varepsilon \in \mathbb{R}\) such that \(0 < 2\varepsilon < F_1(x_1^-) - F_2(x_2)\), we can exploit the right continuity and the monotonicity of \(F_2\) to find \(\delta_1 > 0\) such that \(F_2(\tilde{x}_2) < F_2(x_2) + \varepsilon\) for all \(\tilde{x}_2 \in (x_2 - \delta_1, x_2 + \delta_1)\). Similarly, we can use the monotonicity of \(F_1\) and the definition of left limit to find \(\delta_2 > 0\) such that \(F_1(\tilde{x}_1) < F_1(x_1^-) - \varepsilon\) for all \(\tilde{x}_1 \in (x_1 - \delta_2, x_1 + \delta_2)\). Setting \(\delta = \min\{\delta_i : i = 1, 2\}\), we obtain \(F_2(\tilde{x}_2) < F_2(x_2) + \varepsilon < F_1(x_1^-) - \varepsilon < F_1(\tilde{x}_1)\) for every \((\tilde{x}_1, \tilde{x}_2) \in (x_1 - \delta, x_1 + \delta) \times (x_2 - \delta, x_2 + \delta)\). So, \(H(\tilde{x}_1, \tilde{x}_2) = \min\{F_1(\tilde{x}_1), F_2(\tilde{x}_2)\} = F_2(\tilde{x}_2)\) and hence \(H\) does not depend on \(\tilde{x}_1\) in a small square \(S_1\) around \((x_1, x_2)\). Now setting \(A_1 = (x_1 - \delta, x_2 - \delta), A_2 = (x_1 - \delta, x_2 + \delta), A_3 = (x_1 + \delta, x_2 + \delta)\) and \(A_4 = (x_1 + \delta, x_2 - \delta)\), and using the definition of rectangles given in (4.6), we have \(\gamma(S_1) = \sum_{i=1}^4 (-i)^{i+1} \gamma(R_{A_i}) = \sum_{i=1}^4 (-i)^{i+1} H(A_i) = 0\), since \(H\) does not depend on \(x_1\) on \(S_1\). This proves that \((x_1, x_2) \neq \text{spt}(\gamma)\).

(c) Now we claim that \(\text{spt}(\gamma)\) is a monotone subset of \(\mathbb{R}^2\). Let \((x_1, x_2), (\tilde{x}_1, \tilde{x}_2)\) in \(\text{spt}(\gamma)\) and assume \(x_1 > \tilde{x}_1\), being the opposite case analogous.

We have to prove \(x_2 \geq \tilde{x}_2\). Since \(F_1\) is nondecreasing and (4.9) holds, we have \(F_2(x_2) \geq
\( F_1(x_1) \geq F_1(\tilde{x}_1) \geq F_2(\tilde{x}_2) \). If \( F_2(x_2) > F_2(\tilde{x}_2) \), could conclude because also \( F_2 \) is non-decreasing. Otherwise, all the quantities are equal, \( F_2(x_2) = F_1(x_1) = F_1(\tilde{x}_1) = F_2(\tilde{x}_2) \). Assume by contradiction that \( \tilde{x}_2 > x_2 \). This means that \( F_1 = F_2 = \text{const.} \) on \([x_1, \tilde{x}_1]\) and \([x_2, \tilde{x}_2]\) respectively. Let us show that this would imply that \((x_1, x_2), (\tilde{x}_1, \tilde{x}_2) \notin \text{spt}(\gamma)\). Consider \((\tilde{x}_1, \tilde{x}_2)\), being the proof similar for \((x_1, x_2)\). Using monotonicity of \( F_i \), \( i = 1, 2 \), their right continuity and the definition of left limit, one can find as in (b) \( \delta > 0 \) small enough such that \( H(\tilde{x}_1, \tilde{x}_2) \) does not depend on \( x_2 \) in the square \( S_2 \) with vertices \((\tilde{A}_1, \tilde{A}_2)\) with \( \tilde{A}_i \in \{\tilde{x}_i - \delta, \tilde{x}_i + \delta\}, i = 1, 2 \). This implies \( \gamma(S_2) = 0 \) and therefore \((\tilde{x}_1, \tilde{x}_2) \notin \text{spt}(\gamma)\).

(d) Since \( \text{spt}(\gamma) \) is a monotone subset of \( \mathbb{R} \), the subdifferential of a convex function on \( \mathbb{R} \), \( \gamma \) has to be optimal transference plan by Theorem 3.19.

(e) It remains to prove (4.8). We claim that

\[
\gamma = (F_1^{-1} \times F_2^{-1}) \# \mathcal{L}^1
\]  

(4.10)

Of course, it is enough to prove it on rectangles \( R(x_1, x_2) \) for every \( x_i \), \( i = 1, 2 \), for all \( x_i \in \mathbb{R} \) with \( i = 1, 2 \), \( \gamma(R(x_1, x_2)) = \mathcal{L}^1(\{y \in [0, 1] : (F_1^{-1}(y), F_2^{-1}(y)) \in R(x_1, x_2)\}) \). Notice that the subset of \([0, 1]\) in the right hand side, is \( \bigcap_{i=1,2} \{y \in [0, 1] : F_i^{-1}(y) \leq x_i\} \), and these sets coincides with either \([0, F_i(x_i)]\) or \([0, F_i(x_i)]\), \( i = 1, 2 \), respectively. In any case, the intersection will result an interval with endpoints 0 and \( \min\{F_i(x_i) : i = 1, 2\} \), whose \( \mathcal{L}^1 \) measure is \( \min\{F_i(x_i) : i = 1, 2\} = H(x_1, x_2) \). This proves our claim.

(f) We can now conclude, since (4.10) implies that for any nonnegative measurable function \( \zeta \) on \( \mathbb{R}^2 \) we have

\[
\int_{\mathbb{R}^2} \zeta(x_1, x_2) d\gamma(x_1, x_2) = \int_0^1 \zeta(F_1^{-1}(y), F_2^{-1}(y)) dy.
\]

Choosing \( \zeta = c \), we obtain (4.8). \( \square \)

**Remark 4.20** In fact, the same \( \gamma \) given in Theorem 4.19 is optimal not only for the quadratic distance, but for any cost function \( c(x_1, x_2) \) of the form \( c(x_1 - x_2) \) with \( c : \mathbb{R} \to \mathbb{R} \) nonnegative, convex and symmetric. In this slightly more general case, see Proposition 4.37 and the optimal transportation cost takes the expression (4.14). In the case of a concave cost
function, the situation is completely different: see for example Chapter 2 of [115].

4.1.3 Monotone transport maps and plans on $\mathbb{R}$

We have seen in Section 3.3 that the optimal transport map for the quadratic cost function in the higher dimensional case ($d > 1$) is well–defined as soon as the starting measure has no atoms and is monotone nondecreasing. More precisely, if one uses the sufficient optimality condition given in Theorems 3.19, then one can obtain a uniqueness result for the optimal transport. If additionally $\mu_1$ does not give mass to points, then a canonical map $T : \mathbb{R} \rightarrow \mathbb{R}$ given by $T(x) = F_2^{-1} \circ F_1(x)$ for $x \in \mathbb{R}$ (see Theorem 4.24 below), transports $\mu_1$ onto $\mu_2$, and solves the Monge problem (MP). Indeed, since $F_1$ is continuous, it pushes forward $\mu_1$ to the uniform measure on $(0, 1)$, while the inverse function $F_2^{-1}$ pushes forward the uniform measure to $\mu_2$. In other words, if $\mu_1$ is a given nonatomic probability measure on the Borel sets of $\mathbb{R}$ with a (continuous) distribution function $F_1$, any other probability measure $\mu_2$ on the real line with the distribution function $F_2$ can be obtained as a monotone transform of $\mu_1$, as the distribution of a monotone map $T$ under $\mu_2$. One then says that $T$ pushes forward $\mu_1$ to $\mu_2$ (or $F_1$ to $F_2$) and writes in symbols $T\#\mu_1 = \mu_2$. In particular, notice that we have

$$\int_{-\infty}^{x} d\mu_1 = \int_{-\infty}^{T(x)} d\mu_2.$$  \tag{4.11}

This expresses the general fact that the solution to the transportation problem is given by the monotone rearrangement of $\mu_1$ onto $\mu_2$ (one proceeds to transfer the sand into the hole starting from the left), optimal plans are so called monotone rearrangements. Indeed, any monotone map $T$ sending an atomless $\mu_1$ onto $\mu_2$ should be optimal for the quadratic cost by Theorem 3.19; then, by applying Remark 3.33 and Theorem 3.19, we get the uniqueness of the optimizer, and hence the uniqueness of the monotone map transporting $\mu_1$ onto $\mu_2$.

Before starting to construct the monotone transport map let us take a look at some properties of the generalized inverse of c.d.f..

**Proposition 4.21** If $\mu \in \mathcal{P}(\mathbb{R})$ and $F^{-1}$ is the generalized inverse of its c.d.f. $F$, then $F^{-1}_\#\chi = \mu$. Moreover, given $\mu_i \in \mathcal{P}(\mathbb{R})$, $i = 1, 2$, if we set $\gamma := (F_1^{-1}, F_2^{-1})_\#\chi$, then $\gamma \in \Gamma(\mu_1, \mu_2)$ and $\gamma((\infty, x_1] \times (\infty, x_2]) = F_1(x_1) \wedge F_2(x_2)$.  


Proof. For the first part of the statement, using the proof of Proposition 4.4, which proves that the image measure is $\mu$ by using the characterization of a measure through its c.d.f. (see (4.1)). For the second part of the statement, $\gamma \in \Gamma(\mu_1, \mu_2)$ is just a consequence of the first. Then, let us compute

$$\gamma((\mathbb{R}, x_1] 	imes (\mathbb{R}, x_2]) = \chi\{y \in [0, 1] : F_1^{-1}(y) \leq x_1, F_2^{-1}(y) \leq x_2\}$$

$$= \chi\{y \in [0, 1] : F(x_1) \geq y, F(x_2) \geq y\}$$

$$= F_1(x_1) \wedge F_2(x_2),$$

which is the desired equality. \qed

Definition 4.22 The transport plan $\gamma := (F_1^{-1}, F_2^{-1})_\# \chi$ is said to be the co–monotone transport plan between $\mu_1$ and $\mu_2$ and we denote it by $\gamma_{\text{mon}}$.

Now, consider two measures $\mu_i \in \mathcal{P}(\mathbb{R}), i = 1, 2$: we want to build a monotone transport map (and not only a plan) sending $\mu_1$ onto $\mu_2$, provided $\mu_1$ is a nonatomic. We first need a simple lemma on the c.d.f. of nonatomic measures.

Lemma 4.23 If $\mu_1 \in \mathcal{P}(\mathbb{R})$ is a nonatomic, then $F_\# \mu = \chi$. As a consequence, for every $l \in [0, 1]$, the set $\{x : F(x) = l\}$ is $\mu$–negligible.

Proof. First note that $F$ is continuous because $\mu$ is a nonatomic. Hence, for $a \in [0, 1]$, the set $\{x : F(x) \leq a\}$ is a closed interval of the form $(-\infty, x_a]$, with $F(x_a) = a$. Hence, $F_\# \mu(\mathbb{R} \cap (-\infty, a]) = \mu(\{x : F(x) \leq a\}) = F(x_a) = a$, which proves $F_\# \mu = \chi$. As a consequence, for $l \in [0, 1]$, the sets $\{x : F(x) = l\}$ are $\mu$–negligible, since otherwise the image measure $F_\# \mu$ would have an atom at $l$, which contradicts the first part of the statement. \qed

Theorem 4.24 Given $\mu_i \in \mathcal{P}(\mathbb{R}), i = 1, 2$, suppose that $\mu_1$ is a nonatomic. Then, there exists a unique (monotone) nondecreasing map $T := F_2^{-1} \circ F_1 : \mathbb{R} \to \mathbb{R}$ such that $T_\# \mu_1 = \mu_2$.

Proof. First, let us build one such a map. Let us consider the c.d.f. $F_i, i = 1, 2$ and define the map $T$ through $T(x) := F_2^{-1}(F_1(x)) = \min\{x \in \mathbb{R} : F_1(x) \leq F_2(x)\}$. This quantity is well–defined and belongs to $\mathbb{R}$ provided $F_1(x) \in (0, 1)$ (so that the set on which we take the infimum in Definition 4.1 is neither empty nor unbounded from below). To prove that the above minimum is well–defined, we just use the fact that $F_2$ is right–continuous, so that if
a sequence \( x_k \to x^+ \) as \( k \to +\infty \) satisfies \( F_2(x_k) \geq c \) for a constant \( c \), then we also have \( F_2(x) \geq c \). Hence, the minimum is attained, provided the set of \( x \) satisfying \( F_2(x) \geq c \) is bounded from below, which is true for \( c > 0 \) (since \( \lim_{x \to -\infty} F_2(x) = 0 \)). Since for \( c = 0 \) this does not work, the function \( T \) is well-defined for every \( x \) such that \( F_1(x) > 0 \). The sets \( A_1 = \{ x : F_1(x) = 0 \} \) and \( A_2 = \{ x : F_1(x) = 1 \} \) are \( \mu_1 \)-negligible thanks to Lemma 4.23. Indeed, the sets \( A_1, A_2 \) are intervals of the form \( (-\infty, a] \) (these intervals are closed since \( F_1 \) is continuous, as a consequence of \( \mu_1 \) being a nonatomic, and are bounded from above since \( \lim_{x \to +\infty} F_1(x) = 1 \)). Moreover, we have \( \mu_1(A_1) = \mu_1(A_2) = 0 \), by definition of \( F_1 \). Hence, \( T \) is well-defined \( \mu_1 \)-a.e.

The fact that \( T \) is monotone nondecreasing is obvious; we just have to prove \( T\#\mu_1 = \mu_2 \). Since we already know \( (F_2^{-1})\#\chi = \mu_2 \), by composition we just need to use Lemma 4.23, which proves \( (F_1)\#\mu_1 = \chi \). To check this, it is enough to prove \( \mu_2(-\infty, b] = \mu_1(T^{-1}((-\infty, b])) \) for every \( b \). The condition \( T(x_1) \leq b \) means \( F_2(b) \geq F_1(x) \). Suppose \( F_2(b) < 1 \): in this case the set \( B = T^{-1}(-\infty, b] \) is an interval of the form \( (-\infty, d] \) (since \( F_1 \) is continuous and \( B \) must be bounded), and \( F_1(d) = F_2(b) \). This implies \( \mu_1(B) = F_2(b) \) which is the desired result. On the other hand, if \( F_2(b) = 1 \), then the set \( B = T^{-1}(-\infty, b] \) coincides with \( \mathbb{R} \) and one has \( \mu_1(B) = 1 \), which concludes the proof.

We now pass to the proof of uniqueness. Consider any nondecreasing map \( T \) such that \( T\#\mu_1 = \mu_2 \). From monotonicity we have \( T^{-1}((-\infty, T(x)]) \supset (-\infty, x] \). We deduce \( F_1(x) = \mu_1((-\infty, x]) \leq \mu_1(T^{-1}((-\infty, T(x)])) = \mu_2((-\infty, T(x)]) = F_2(T(x)) \), which means \( T(x) \geq F_2^{-1}(F_1(x)) \). Suppose that the inequality is strict. This means that there exists \( \varepsilon_0 > 0 \) such that \( F_2(T(x) - \varepsilon) \leq F_1(x) \) for every \( \varepsilon \in (0, \varepsilon_0) \). On the other hand, from \( T^{-1}((-\infty, T(x) - \varepsilon)) \subset (-\infty, x) \), we get \( F_2(T(x) - \varepsilon) \leq F_1(x) \). Hence, we obtain \( F_2(T(x) - \varepsilon) = F_1(x) \) for every \( \varepsilon \in (0, \varepsilon_0) \). Thus, \( F_2(T(x) - \varepsilon) \) is the value that \( F_2 \) takes on an interval where it is constant. In other words, the points where different values of \( T \) are allowed are the preimages through \( F_1 \) of flat parts of \( F_2 \). These intervals of positive length where \( F_2 \) is flat are a countable quantity; hence the values of \( F_2 \) on those intervals are also countable. We denote by \( l_i \) these values and \( L = \bigcup_i \{ x : F_1(x) = l_i \} \). As a consequence, the points \( x \) where \( T(x) > F_2^{-1}(F_1(x)) \) are contained in \( L \). Lemma 4.23 proves that the set \( L \) is negligible. Indeed, for each value \( l_i \in \mathbb{R} \), we have \( \mu_1(F_1^{-1}(\{l_i\})) = 0 \), since the preimage of a point is
an interval, and the fact that $F_1$ is constant on an interval means that there is no mass on it (up to, possibly, an atom at the first point, which is not possible for $\mu_1$). This means that $\mu_1(F_1^{-1}(L)) = 0$, which shows that, outside a negligible set, the value of the map $T$ is prescribed, and gives uniqueness. This allows us to conclude $T(x) = F_2^{-1}(F_1(x)) \mu_1$-a.e. 

Notice that we just proved the existence, and then the uniqueness of a monotone map $T$ with given image measure. We sometimes call this map $T_{\text{mon}}$.

**Remark 4.25** The latter proof was complicated by the possibility that the c.d.f. could be either discontinous (that discontinuity points of $F_2$ correspond to atoms for $\mu_2$) or not strictly increasing. Whenever $\mu_2$ has an atom (or, a jump occurs in $F_2$), we have an interval on which $T_{\text{mon}} = F_2^{-1} \circ F_1$ is constant, there is a point in which we have to keep putting mass before moving on. Should $F_2$ be continuous and strictly monotone (which means that $\mu_2$ is a nonatomic and supported on the whole $\mathbb{R}$), then one would simply have the monotone nondecreasing map $T_{\text{mon}}$.

**Remark 4.26** As a consequence of the explicit formula of the previous remark, one can also study the regularity of this map $T_{\text{mon}}$ depending on the regularity of $\mu_i, i = 1, 2$. Indeed, as soon as these measures are fully supported and have no atoms, the two functions $F_i, i = 1, 2$ are homeomorphisms, and so is $T_{\text{mon}}$. Moreover, if $\mu_i, i = 1, 2$ are absolutely continuous, with continuous densities which do not vanish, then they are also diffeomorphisms, and so is $T_{\text{mon}}$. In general, the regularity of $T_{\text{mon}}$ is one degree higher than that of the two measures, as it is the case for $F_i, F_i^{-1}, i = 1, 2$.

In the next subsection we will see that the map $T_{\text{mon}}$ that we have just built optimizes a whole class of transport costs. To prove it, we will need the following characterization of $\gamma_{\text{mon}}$ and $T_{\text{mon}}$.

**Lemma 4.27** Let $\gamma \in \Gamma(\mu_1, \mu_2)$ be a transport plan between two measures $\mu_i \in \mathcal{P}(\mathbb{R}), i = 1, 2$. Suppose that it satisfies the property

$$(x_1, x_2), (\tilde{x}_1, \tilde{x}_2) \in \text{spt}(\gamma), x_1 < \tilde{x}_1 \implies x_2 < \tilde{x}_2. \quad (4.12)$$
Then, we have \( \gamma = \gamma_{\text{mon}} \). In particular there is a unique \( \gamma \) satisfying (4.12). Moreover, if \( \mu_1 \) is a nonatomic, then \( \gamma = \gamma_{T_{\text{mon}}} \), is induced by a transport map, and, more precisely, by \( T_{\text{mon}} \). In particular one only transport plan can satisfy this condition, provided \( \mu_i, i = 1, 2 \) are fixed and \( \mu_1 \) has no atoms.

**Proof.** For the first part of the statement, we just need to prove \( \gamma((-\infty, x_1] \times (-\infty, x_2]) = F_1(x_1) \land F_2(x_2) \). Indeed, from Proposition 4.21, this condition is satisfied by \( \gamma_{\text{mon}} \), and we saw that this is enough to characterize a measure on \( \mathbb{R}^2 \). Consider the two sets \( A = (-\infty, x_1] \times (x_2, +\infty) \) and \( B = (x_1, +\infty) \times (-\infty, x_2] \). From the assumption on \( \gamma \), it is not possible to have both \( \gamma(A) > 0 \) and \( \gamma(B) > 0 \) (otherwise we would have points in \( \text{spt}(\gamma) \) violating condition (4.12)). Hence, we can write

\[
\gamma((-\infty, x_1] \times (-\infty, x_2]) = \gamma((-\infty, x_1] \times (-\infty, x_2) \cup A) \land \gamma((-\infty, x_1] \times (-\infty, x_2) \cup B)
\]

\[
= \gamma((-\infty, x_1] \times \mathbb{R}) \land \gamma(\mathbb{R} \times (-\infty, x_2]) = F_1(x_1) \land F_2(x_2).
\]

This proves the first part of the claim.

For the second part, we suppose \( \mu_1 \) to be atomless. For any point \( x \in \mathbb{R} \) one can define the interval \( I_x \) as the minimal interval \( I \) such that \( \text{spt}(\gamma) \cap \{x\} \times \mathbb{R} \subset \{x\} \times I \). This interval can obviously be reduced to a singleton. The assumption on \( \gamma \) implies that the interior of all these intervals is disjoint (and ordered). In particular there can be at most a countable quantity of points such that \( I_x \) is not a singleton. Since these points are \( \mu_1 \)-negligible (as a consequence of \( \mu_1 \) being nonatomic), we can define \( \mu_1 \)-a.e. a map \( T \) such that \( \gamma \) is concentrated on the graph of \( T \). This map will be monotone nondecreasing because of the implication (4.12) above, and this gives \( T = T_{\text{mon}} \) since we already know the uniqueness of a nondecreasing map with fixed marginals (Theorem 4.24).

Notice that, thanks to Remark 4.26, it is interesting to investigate the regularity of \( T_{\text{mon}} \) according to the regularity of the densities of \( \mu_i, i = 1, 2 \). If we suppose that \( \mu_1 \) has no atoms and that \( \mu_2 \) is supported on a whole interval \( [a, b] \subset \mathbb{R} \) we get the equality \( F_2 \circ T_{\text{mon}} = F_1 \).

If \( \mu_i = \rho_i(x_i) \, dx_i, i = 1, 2 \) are absolutely continuous, then \( F_1 \) has one degree of regularity more than \( \rho_1 \) and \( F_2 \) one more than \( \rho_2 \). Moreover, if \( \rho_2 \neq 0 \) then \( F_2' \neq 0 \). This means that \( T_{\text{mon}} \) has one regularity more than the worse between \( \rho_1 \) and \( \rho_2 \), provided \( \rho_2 \neq 0 \).
Remark 4.28 If $\mu_i \ll L^1$, $i = 1, 2$ with continuous densities $\rho_i$, respectively, and with $\rho_2 > 0$, then the optimal plan for Monge problem is of class $C^1$ and, deriving with respect to $x$ in (4.11), we obtain that $T$ satisfies the Monge–Ampère PDE equation (3.24) in the one-dimensional Euclidean space; that is, $\rho_1(x) = \rho_2(T(x)) T'(x)$.

4.1.4 The optimality of monotone maps for a convex cost

Now that we know quite well the properties, definitions, and characterizations of the map $T_{\text{mon}}$ and of the plan $\gamma_{\text{mon}}$, we can see that they are, in the 1D case, optimal for several different costs and not only for the quadratic one. This is really specific to the 1D case; it will not be true in higher dimension. The costs that we will consider will be convex functions of the difference $x_1 - x_2$ and, to stress the possible asymmetric behavior of these costs, we prefer to write $c(x_2 - x_1)$ instead of $c(x_1 - x_2)$.

Theorem 4.29 Let $c : \mathbb{R} \to \mathbb{R}^+$ be a strictly convex function, and $\mu_i \in \mathcal{P}(\mathbb{R})$, $i = 1, 2$ be probability measures. Consider the cost $c(x_1, x_2) = c(x_2 - x_1)$ and suppose that the optimal transport problem (KP) has a finite value. Then, (KP) has a unique solution, which is given by $\gamma_{\text{mon}}$. In the case where $\mu_1$ is nonatomic, this optimal plan is induced by the map $T_{\text{mon}}$. Moreover, if the strict convexity assumption is withdrawn and $c$ is only convex, then the same $\gamma_{\text{mon}}$ is actually an optimal plan, but no uniqueness is guaranteed anymore.

Proof. We will use the fact that the support of any optimal $\gamma$ is a $c$–cyclically monotone set $\mathcal{G}$. This means in particular that $(x_1, x_2), (\tilde{x}_1, \tilde{x}_2) \in \mathcal{G}$ implies

$$c(x_2 - x_1) + c(\tilde{x}_2 - \tilde{x}_1) \leq c(\tilde{x}_2 - x_1) + c(x_2 - \tilde{x}_1). \quad (4.13)$$

We only need to show that this implies (in the strictly convex case) a monotone behavior: we will actually deduce from (4.13) that $x_1 < \tilde{x}_1$ implies $x_2 \leq \tilde{x}_2$, and this will allow to conclude as we noticed above.

To prove $x_2 \leq \tilde{x}_2$, we suppose by contradiction $x_2 > \tilde{x}_2$ and denote $a_1 = x_2 - x_1$, $a_2 = \tilde{x}_2 - \tilde{x}_1$ and $\delta = \tilde{x}_1 - x_1$. Condition (4.13) reads $c(a_1) + c(a_2) \leq c(a_2 + \delta) + c(a_1 - \delta)$. Moreover, the assumption $\tilde{x}_2 < x_2$ implies $a_2 + \delta < a_1$. We also need to recall that $\delta > 0$ ($x_1 < \tilde{x}_1$): this
implies that $a_2 + \delta$ and $a_1 - \delta$ are locates in the segment between $a_2$ and $a_1$ (and $a_2 < a_1$). More precisely, we have $a_2 + \delta = \lambda a_1 + (1 - \lambda) a_2$ and $a_1 - \delta = (1 - \lambda) a_1 + \lambda a_2$, for $0 < \lambda < 1$ where $\lambda := \delta/(a_1 - a_2)$. Thus, strict convexity yields $c(a_1) + c(a_2) \leq c(a_2 + \delta) + c(a_1 - \delta) < \lambda c(a_1) + (1 - \lambda)c(a_2) + (1 - \lambda)c(a_1) + \lambda c(a_2) = c(a_1) + c(a_2)$. This gives a contradiction and proves the thesis in the strictly convex case.

The statement when $c$ is only convex is trivial if $c$ is constant (since every $\gamma$ is optimal), and, if not, it is obtained by approximation. Lemma 4.30 below proves that there exists a sequence of strictly convex functions $c_\varepsilon$ such that $c \leq c_\varepsilon \leq (1 + \varepsilon)c + \varepsilon$. Let us take the transport cost $c_\varepsilon(x_1, x_2) = c_\varepsilon(x_2 - x_1)$. In this case we know that $\gamma_{mon}$ optimizes the cost $\int c_\varepsilon d\gamma$ and hence $\int c(x_2 - x_1) d\gamma_{mon} \leq \int c_\varepsilon(x_2 - x_1) d\gamma_{mon} \leq \int c_\varepsilon(x_2 - x_1) d\gamma \leq \varepsilon + (1 + \varepsilon) \int c(x_2 - x_1) d\gamma$ for all $\gamma \in \Gamma(\mu_1, \mu_2)$. Passing to the limit as $\varepsilon \to 0$, we get that $\gamma_{mon}$ also optimizes the cost $c$.

**Lemma 4.30** For every nonconstant convex and positive function $c : \mathbb{R} \to \mathbb{R}^+$ and every $\varepsilon > 0$, there exists $c_\varepsilon : \mathbb{R} \to \mathbb{R}^+$ strictly convex, such that $c \leq c_\varepsilon \leq (1 + \varepsilon)c + \varepsilon$.

**Proof.** We just need to prove that there is a strictly convex function $f : \mathbb{R} \to \mathbb{R}^+$ such that $f \leq c + 1$. Then, we will take $c_\varepsilon := c + \varepsilon f$. From the fact that $c$ is convex, it is bounded from below by an affine function, and we have $c(t) \geq (x_1 t + x_2)^+$ (we take the positive part since we also know $c \geq 0$). It can be checked that $f(t) := (\sqrt{4 + (x_1 t + x_2)^2} + (x_1 t + x_2))/2$ is strictly convex and satisfies $f(t) \leq (2 + |x_1 t + x_2|+(x_1 t + x_2))/2 = 1 + (x_1 t + x_2)^+$. □

**Remark 4.31** Positivity of the function $c$ is not really necessary, as soon as $\mu_i$, $i = 1, 2$ satisfy some integrability conditions so that $\inf (K') > -\infty$. For instance, if we have $\int |x_i| d\mu_i(x_i) < +\infty$, $i = 1, 2$, we can add an affine function to $c$ and make it convex, and the cost of this affine function does not depend on $\gamma$ and is finite (see Example 2.14 below).

**Remark 4.32** We stress that a completely analogous proof could be used to prove that optimal maps for strictly concave costs (of the form $c(x_1 - x_2)$ for $c$ strictly concave) are monotone decreasing instead of monotone increasing. Compare with what is described in Section 3.3.2 [104].
Remark 4.33 Also, the optimality of $\gamma_{\text{mon}}$ is true under more general assumptions; that is, for some costs which are not of the form $c(x_1 - x_2)$ but satisfy the twist condition. See Exercise 10 of [104].

Easy examples where $c$ is not strictly convex and $T_{\text{mon}}$ is not the unique optimal transport map may be built as follows.

Example 4.34 (Linear costs). Suppose that $c(x_1, x_2) = L(x_1 - x_2)$, the map $a : \mathbb{R}^d \to \mathbb{R}$ being linear. In this case any transport plan $\gamma$ is optimal, and any transport map as well. This can be easily seen if one writes $\int L(x_1 - x_2) \, d\gamma = \int L(x_1) \, d\gamma - \int L(x_2) \, d\gamma = \int L(x_1) \, d\mu_1 - \int L(x_2) \, d\mu_2$, which shows that the result does not depend on $\gamma$ but only on its marginals. This general example works for $\mu_1, \mu_2$ compactly supported (so that we do not have any problem of integrability of $L(x_1)$ and $L(x_2)$), and in any dimension. Hence, also in 1D.

Example 4.35 (Distance cost on the line). Suppose that $c(x_1, x_2) = |x_1 - x_2|$ and that $\mu_i \in \mathcal{P}(\mathbb{R}), i = 1, 2$ are such that $\sup \text{spt}(\mu_1) < \inf \text{spt}(\mu_2)$. In this case also any transport plan $\gamma$ is optimal, and any transport map as well. This can be seen by noticing that for every $(x_1, x_2) \in \text{spt}(\mu_1) \times \text{spt}(\mu_2)$ we have $c(x_1, x_2) = x_2 - x_1$, which is again a linear cost.

Example 4.36 (Book shifting). Consider $c(x_1, x_2) = |x_1 - x_2|$, $\mu_1 = \frac{1}{2} \mathcal{L}^1|_{[0,2]}$ and $\mu_2 = \frac{1}{3} \mathcal{L}^1|_{[1,3]}$. Then $T_{\text{mon}}(x_1) = x_1 + 1$ is the monotone transport plan transporting $\mu_1$ onto $\mu_2$. Its cost is $T_{\text{mon}}(\mu_1, \mu_2) = \int |T_{\text{mon}}(x_1) - x_1| \, d\mu_1 = 1$. Yet, the transport map $T$ given by $T(x_1) = x_1 + 2$ if $x_1 \leq 1$ and $T(x_1) = x_1$ if $x > 1$ also satisfies $T_{\#} \mu_1 = \mu_2$ and $\int |T(x_1) - x_1| \, d\mu_1 = \frac{1}{2} \int_0^1 2 \, dx_1 = 1$, and is optimal as well.

Starting from the fact that the optimal transport for all these costs is the monotone one, we can express the cost for sending a given measure $\mu_1$ onto another measure $\mu_2$ in terms of their c.d.f. $F_1$ and $F_2$. We will do it in the easier case where both c.d.f. are continuous and strictly increasing (that is, the two measures $\mu_1$ and $\mu_2$ have no atoms and have full supports on $\mathbb{R}$).
Proposition 4.37  Given \( \mu_i \in \mathcal{P}(\mathbb{R}) \), \( i = 1, 2 \), consider the cost function \( c(x_1, x_2) = c(x_1 - x_2) \), where \( c \) is a convex function. Then, we have the total transportation cost associated with it can be given by

\[
C(\mu_1, \mu_2) = \int_0^1 c(F_2^{-1}(y) - F_1^{-1}(y)) \, d\mathcal{L}^1(y),
\]

where \( F_i, \, i = 1, 2 \) are the c.d.f. associated with the probability measures \( \mu_i \); respectively, from the space \( \mathcal{P}(\mathbb{R}) \). In particular, if \( c(z) = |z| \), then this also coincides with \( \int_\mathbb{R} |F_1(x) - F_2(x)| \, dx \).

\textbf{Proof.} The first part of the statement is just a consequence of the optimality of \( \gamma_{\text{mon}} \) (Theorem 4.29). We consider the optimal transport problem for the cost \( c(x_1, x_2) = c(x_1 - x_2) \) from \( \mu_1 \) to \( \mu_2 \): we know that \( T = F_2^{-1} \circ F_1 \) is optimal, and hence the optimal cost equals

\[
\int_\mathbb{R} c(F_2^{-1}(F_1(x)) - x) \, d\mu_1(x) = \int_0^1 c(F_2^{-1}(y) - F_1^{-1}(y)) \, dy,
\]

where the last equality comes from the fact that \( \text{Id} = F_1 \circ F_1^{-1} \) and from \( F_1 \# \mu_1 = \chi \).

The particular case of \( c(x_1, x_2) = |x_1 - x_2| \) may be treated by geometric consideration: indeed, the integral \( \int_0^1 |F_2^{-1}(y) - F_1^{-1}(y)| \, dy \) equals the area of the part of the strip \( [0, 1] \times \mathbb{R} \) bounded by the graphs of \( F_1^{-1} \) and \( F_2^{-1} \). In order to pass from the inverse functions to the direct ones it is enough to turn the head and look at the same strip from the variable \( x \) instead of \( y \). But, if we want to prove it through computations, we have

\[
\int |F_2^{-1}(y) - F_1^{-1}(y)| \, dy = \mathcal{L}^2(\{(y, x) \in [0, 1] \times \mathbb{R} : F_1^{-1}(y) < x < F_2^{-1}(y) \text{ or } F_2^{-1}(y) < x < F_1^{-1}(y)\})
\]

\[
+ \mathcal{L}^2(\{(y, x) \in [0, 1] \times \mathbb{R} : F_2^{-1}(y) < x < F_1^{-1}(y)\} \cup \{(y, x) \in [0, 1] \times \mathbb{R} : F_1^{-1}(y) < x < F_2^{-1}(y)\})).
\]

Then, by Fubini’s theorem, we have

\[
\mathcal{L}^2(\{(y, x) \in [0, 1] \times \mathbb{R} : F_1^{-1}(y) < x < F_2^{-1}(y)\}) = \int_\mathbb{R} \mathcal{L}^1(\{y \in [0, 1] : y < F_1(x) \text{ and } F_2(x) < y\}) \, dt = \int_\mathbb{R} \mathcal{L}^1(\{y \in [0, 1] : F_2(x) < y < F_1(x)\}) \, dx.
\]
Analogously,
\[ \mathcal{L}^2(\{(y, x) \in [0, 1] \times \mathbb{R} : F_2^{-1}(y) < x < F_1^{-1}(y)\}) = \int_{\mathbb{R}} \mathcal{L}^1(\{y \in [0, 1] : F_1(y) < y < F_2(x)\}) \, dx, \]
and, summing up,
\[ \mathcal{L}^2(\{(y, x) \in [0, 1] \times \mathbb{R} : F_1^{-1}(y) < x < F_2^{-1}(y)\}) = \int_{\mathbb{R}} \mathcal{L}^1(\{y \in [0, 1] : F_2(y) < y < F_1(x) \text{ or } F_1(x) < y < F_2(x)\}) \, dx = \int_{\mathbb{R}} |F_1(x) - F_2(x)| \, dx. \]
This concludes the proof. \(\square\)

Notice that the computation of the areas in the last part allows (graphically) to seeing the integral of \(|F_2^{-1}(y) - F_1^{-1}(y)|\) corresponds to “horizontally”, while if seen “vertically” we get the integral of \(|F_1(x) - F_2(x)|\). In this case, the optimal transportation cost coincides with the \(L^1\) distance between c.d.f. which will be presented in the next section.

Finally, under the weakened assumption (3.16) for optimality, we can prove the same result of Theorem 4.29, concerning necessary and sufficient optimality conditions and the existence of optimal transport in (KP).

**Proposition 4.38** Let \(c \in C^1(\mathbb{R})\) be a nonnegative, strictly convex function. Let \(\mu_i, i = 1, 2\) be two Borel probability measures on \(\mathbb{R}\) satisfying
\[ \iint_{\mathbb{R}^2} c(x_1, x_2) \, d\mu_1(x_1) \, d\mu_2(x_2) < +\infty \quad (4.15) \]
If \(\mu_1\) has no atom, and \(F_i\) stand for the respective cumulative distribution of \(\mu_i\), respectively, then \(T := F_2^{-1} \circ F_1\) solves Monge’s problem for the cost \(c(x_1, x_2) = c(x_1 - x_2)\). If \(\gamma\) is the induced transport plan; that is, \(\gamma := (\text{Id}, T)\#\mu_1\), then is optimal for the Monge–Kantorovich problem (KP).

**Proof.** To begin with, notice that \(T\) is well defined for \(\mu_1\)-a.e. (see Theorem 4.24). Indeed, there might be a problem only when \(F_1(x) = 0\), for \(F_2^{-1}(0) = -\infty\). But \(F_1 \equiv 0\) only on \((-\infty, a]\) for some \(a \in \mathbb{R}\), and, by Definition 4.1 of \(F_1\), we have \(\mu_1([-\infty, a]) = F_1(a) = 0\).
Notice also that, as $F_1$ and $F_2$ are nondecreasing, $T$ must be nondecreasing as well. Then, Lemma 4.3 above applied to the cumulative distribution $F_2$ yields

$$T^{-1}((-\infty, x_2]) = \{x_1 \in \mathbb{R} : F_2^{-1}(F_1(x_1)) \leq x_2\} = \{x_1 \in \mathbb{R} : F_1(x_1) \leq F_2(x_2)\}.$$

This set has to be an interval, as $T$ is nondecreasing. Since $\mu_1$ has no atom, $F_1$ is increasing and continuous, so this interval must be closed. Thus, if $x_1$ is its supremum, we must have $F_1(x_1) = F_2(x_2)$, and therefore

$$\mu_1(T^{-1}((-\infty, x_2])) = \mu_1((-\infty, x_1]) = F_1(x_1) = F_2(x_2) = \mu_2((-\infty, x_2]).$$

This is enough to show $T_\# \mu_1 = \mu_2$.

Now, let us prove $T$ is optimal. On the one hand, if $\hat{x}_1 \geq x_1$, then, as $T$ and $c'$ are nondecreasing, $c'(\hat{x}_1 - T(\hat{x}_1)) \leq c'(\hat{x}_1 - T(x_1))$. Integrating between $x_1$ and some $x_2 \leq x_1$ we get

$$\int_{x_1}^{x_2} c'(\hat{x}_1 - T(\hat{x}_1)) d\hat{x}_1 \leq \int_{x_1}^{x_2} c'(\hat{x}_1 - T(x_1)) d\hat{x}_1 \leq c(x_2 - T(x_1)) - c(x_1 - T(x_1)).$$

On the other hand, if $\hat{x}_1 \leq x_1$, then $c'(\hat{x}_1 - T(\hat{x}_1)) \geq c'(\hat{x}_1 - T(x_1))$; integrating between $x_1$ and some $x_2 \leq x_1$, we again get

$$\int_{x_1}^{x_2} c'(\hat{x}_1 - T(\hat{x}_1)) d\hat{x}_1 \leq -\int_{x_2}^{x_1} c'(\hat{x}_1 - T(x_1)) d\hat{x}_1 \leq c(x_2 - T(x_1)) - c(x_1 - T(x_1)).$$

Thus, if we set

$$\xi(x_2) := \int_0^{x_2} c'(\hat{x}_1 - T(\hat{x}_1)) d\hat{x}_1,$$

then, in both cases, $\xi(x_2) - \xi(x_1) \leq c(x_2 - T(x_1)) - c(x_1 - T(x_1))$, which implies

$$\xi^c(T(x_1)) := \inf_{x_2} [c(x_2 - T(x_1)) - \xi(x_2)] = c(x_1 - T(x_1)) - \xi(x_1),$$

and this yields $\xi_1$ is $c$–concave. On the other hand, the condition (4.15) ensures that there
are \( \bar{x}_i, i = 1, 2 \) such that
\[
\int_{\mathbb{R}} c(x_1 - \bar{x}_2) \, d\mu_1(x_1) < +\infty \quad \text{and} \quad \int_{\mathbb{R}} c(\bar{x}_1 - x_2) \, d\mu_2(x_2) < +\infty.
\]

Since \( c(x_1 - \bar{x}_2) - \xi^c(\bar{x}_2) \geq \xi(x_1), c(\bar{x}_1 - T(x_1)) - \xi(\bar{x}_1) \geq \xi^c(T(x_1)), \) and \( \xi(x_1) \geq -\xi^c(T(x_1)), \) we have
\[
c(x_1 - \bar{x}_2) - \xi^c(\bar{x}_2) \geq \xi(x_1) \geq -c(\bar{x}_1 - T(x_1)) + \xi(\bar{x}_1)
\]
and as \( T_\# \mu_1 = \mu_2, \) this implies \( \xi \in L^1(\mu_1). \) Similarly, \( \xi^c \in L^1(\mu_2). \) Therefore, integrating the equality \( \xi^c(T(x_1)) - \xi(x_1) = c(x_1 - T(x_1)) \) with respect to \( \mu_1 \) gives
\[
\int_{\mathbb{R}} \xi^c(x_2) \, d\mu_2(x_2) - \int_{\mathbb{R}} \xi(x_1) \, d\mu_1(x_1) = \int_{\mathbb{R}} c(x_1, T(x_1)) \, d\mu_1(x_1).
\]

Since \( \xi^c(x_2) - \xi(x_1) \leq c(x_1 - T(x_1)) \) for all pair \( (x_1, x_2), \) if \( \gamma \) is any other transport plan, the associated total transport cost is necessarily greater, and thus \( T \) is optimal. \( \square \)

### 4.2 Monge–Kantorovich metrics on the real line

A recurrent theme in our work is the so–called Wasserstein distance which may be regarded as the metric side of optimal transportation. We are particularly interested in the Monge–Kantorovich distance on \( \mathbb{R}, \) which is special from the point of view of applications. In this section, we only consider the case of one–dimensional Euclidean space. We refer the reader to Chapter 7 of [17] for the properties of the Wasserstein distance. We will be rather brief and will skip many details. Yet, we will later focus on the \( d = 1 \) case, as this lies at the heart of most of our study. Let us now recall the following proposition

**Proposition 4.39** If \( \mu_i, i = 1, 2 \) are Borel probability measures on the real line and \( \mu_1 \) is atom–free, then it is known by the previous section that there exists a unique (up to a set of \( \mu_1 \)-zero measure) optimal map pushing forward \( \mu_1 \) to \( \mu_2. \) It is called the monotone rearrangement and is obtained as \( F_2^{-1} \circ F_1, \) where \( F_i, i = 1, 2 \) are the c.d.f. of \( \mu_i, \) respectively. We have \( F_2(y) = \mu_2(-\infty, y] \) and \( F_2^{-1}(x) = \{ y \in \mathbb{R} : F_2(y) \geq x \}. \)

Note that \( F_2^{-1} \) is the right–continuous generalized inverse of \( G \) (in [70] the left–continuous
one is considered). In this work, optimal map on the real line always means right continuous optimal map, see Section 4.1 for more details.

4.2.1 The \( p \)-th–Wasserstein space and \( L^p \)-Wasserstein distance

First of all, we give the basic definitions of the \( p \)-th–Wasserstein distance and the \( p \)-th–Wasserstein space on the real line. By definition, \( \mathcal{P}(\mathbb{R}) \) is the set of all Borel probabilities on \( \mathbb{R} \), whereas \( \mathcal{P}_p(\mathbb{R}) \), \( p \geq 1 \) is the collection of all measures in \( \mathcal{P}(\mathbb{R}) \) with finite moments of order \( p \); that is, \( \mu \in \mathcal{P}(\mathbb{R}) \) satisfying \( \int_{\mathbb{R}} |x - x_0|^p d\mu(x) < +\infty \) for some (and hence any) \( x_0 \) in \( \mathbb{R} \). We take \( c_p(x_1, x_2) = |x_1 - x_2|^p \) for some \( p > 0 \). We shall use the notation \( C_p(\mu_1, \mu_2) \) provides by (KP) for the optimal transportation “total” cost between two probabilities \( \mu_1 \) and \( \mu_2 \) on \( \mathbb{R} \) with the cost \( c_p \).

**Definition 4.40** For all \( p > 0 \), the \( p \)-th–Wasserstein distance on the real line is the \( \mathbb{R} \cup \{+\infty\} \)-valued map \( W_p \) defined on \( \mathcal{P}_p(\mathbb{R}) \times \mathcal{P}_p(\mathbb{R}) \) by \( W_p = C_p(\cdot, \cdot) \), where \( s(p) = 1 \) if \( p \leq 1 \) and \( s(p) = 1/p \) if \( p > 1 \). The set \( \mathcal{P}_p(\mathbb{R}) \) endowed with the Wasserstein distance \( W_p \) is called the \( p \)-th–Wasserstein space on the real line.

Notice that this distance is strongly related to the Monge–Kantorovich problem (KP) (originated by Monge [89] and relaxed by Kantorovich [78]—see also [19], [94], [99], [104], [115], [116] for recent surveys). For this, the distance \( W_p \) is commonly known by the Monge–Kantorovich distance of order \( p \) (or, with exponent \( p \)) on \( \mathcal{P}_p(\mathbb{R}) \); that is, the \( p \)-th–root of the minimal transport cost for the cost \( c_p \). In particular, for \( p = 2 \), \( W_2 = T_{c_2}^{1/2} \) is the so–called the quadratic Wasserstein distance.

We also recall the following lemma, whose proof can be found in [115]:

**Lemma 4.41** If \( \mu_i \in \mathcal{P}_p(\mathbb{R}) \), for \( i = 1, 2 \), \( 1 \leq p < +\infty \), and \( \mu_1 \) is atom–free (that is, \( \mu_1 \) does not give mass to small sets ), then the set \( \Gamma_o(\mu_1, \mu_2) \) of where the minimum is achieved is nonempty and reduces to a single element \( \{\gamma_0\} \) (see Remark 3.4). In that case, \( \gamma_0 = (\text{Id} \times \xi')_\#\mu_1 \) for some \( \xi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\} \) convex lower semicontinuous (see Chapter 6 of [17] and [69]). The map \( \xi' \) is the unique (up to a set of \( \mu_1 \)--zero measure) optimal map that such that \( \xi' \) pushing forward \( \mu_1 \) to \( \mu_2 \), and solves the Monge problem (3.1).
What we mean in the previous lemma is that, $\xi'$ is the unique map that minimizes $M \to \int_{\mathbb{R}} |\text{Id} - M|^p \, d\mu_2$ over the set of Borel maps $M$ satisfying $M_\# \mu_1 = \mu_2$. In fact, the definition of $M$ matters only on the support of $\mu_1$.

**Theorem 4.42** Suppose $I := (0, 1) \subset \mathbb{R}$ and let $\chi$ be the restriction of the Lebesgue measure $\mathcal{L}^1$ on $\mathbb{R}$ to the open interval $I$. Then, there is a monotone nondecreasing function $M : I \to \mathbb{R}$ such that $M_t \# \chi = \mu$. It is uniquely defined up to a set of $\chi$–measure and is the optimal map that pushes $\chi$ forward to $\mu_1$.

Notice that, since $\int_{\mathbb{R}} |y|^p \, d\mu_1(y) = \|M_t\|^p_{L^p(I)} < +\infty$ and $M_t$ is monotone, it achieves only finite values in $I$. Hence, the set of discontinuity of $M_t$ is at most countable.

**Remark 4.43** The monotone nondecreasing map $M$ can be described explicitly by the generalized inverse of the c.d.f. of the Borel probability measure (see Section 4.1 for more details).

We next write the expression of the one which is right–continuous and which will be used throughout this work. In fact, it follows that, for any $p \geq 1$, $W_p$ is given by the $L^p$–norm (restricted to $I$) of monotone nondecreasing maps pushing forward $\chi$ to $\mu$ in $\mathcal{P}_p(\mathbb{R})$. The proof of the next corollary is given in Remark 4.45 below.

**Corollary 4.44** For $y \in \mathbb{R}$ and $x \in I$, set

$$M_t(x) = \inf \{ y \in \mathbb{R} : F(y) \geq x \}. \quad (4.16)$$

Then, $M_\# \chi = \mu$ and the $L^p$–Wasserstein distance on $\mathbb{R}$ is given by the well–known identity

$$W_p(\mu_1, \mu_2) = \|M_1 - M_2\|_{L^p(I)}. \quad (4.17)$$

Notice that, the assumption $\mu_i \in \mathcal{P}_p(\mathbb{R})$, $i = 1, 2$ guarantees finiteness of this value since the triangle inequality for $| \cdot |^p$ and the convexity of $z \to z^p$ readily lead to $|x_1 - x_2|^p \leq C (|x_1|^p + |x_2|^p)$ for some $C = \max(1, 2^{p-1})$ and for every $x_i \in \mathbb{R}$, $i = 1, 2$, whence $W_p^1(\mu_1, \mu_2) \leq C (\int |x_1|^p \, d\mu_1(x_1) + \int |x_2|^p \, d\mu_2(x_2)) < +\infty$, where we have taken into account that all $\gamma$’s have marginals $\mu_1$ and $\mu_2$ and their moments of order $p$ are finite. This shows
that $|x_1 - x_2|^p$ is $\gamma$–integrable as soon as $|x|^p$ is $\mu_1$–integrable and $|y|^p$ is $\mu_2$–integrable; that is, the $L^p$–Wasserstein distance $W_p$ is finite on $\mathcal{P}_p(\mathbb{R})$ for $p \geq 0$. We refer the reader to [17], [104], [115], [116] for the properties of $W_p$ and $\Gamma_o(\mu_1, \mu_2)$. In the next remark we comment on how the identity (4.17) may be established.

**Remark 4.45** If $M_0$, $M_T : I \rightarrow \mathbb{R}$ are Borel maps and $M_t\#\chi = \mu_t$ for $t = 0, T$ then $(M_0 \times M_T)\#\chi$ has $\mu_0$ and $\mu_T$ as its marginals and so, $W_p^p(\mu_0, \mu_T) \leq \int_{\mathbb{R}^2} |x - \bar{x}|^p \ d\gamma(x, \bar{x}) = \|M_0 - M_T\|_{L_p(I)}$. Hence, if $M \in AC^p(0, T; L^p(I))$ and $\mu_t := X_t\#\chi$ then $\mu \in AC^p(0, T; \mathcal{P}_p(\mathbb{R}))$. Suppose in addition that $M_0$ and $M_T$ are monotone nondecreasing. For each $k$ integer, we choose $\mu_0^k \in \mathcal{P}_p(\mathbb{R})$ absolutely continuous with respect to $L^1$, of positive density, such that $W_p(\mu_0^k, \mu_0) \leq 1/k$. Let $M_0^k : I \rightarrow \mathbb{R}$ be monotone increasing satisfying $M_0^k\#\chi = \mu_0^k$. The map $M_0^k$ admits left inverse $F^k : \mathbb{R} \rightarrow I$ which is monotone increasing. Since $M_T \circ F^k$ is monotone nondecreasing and pushes $\mu_0^k$ forward to $\mu_T$, we conclude that $W_p^p(\mu_0^k, \mu_T) = \|Id - M_T \circ F^k\|_{L_p(\mu_0^k)}$. The last term is checked to be $\|M_0^k - M_T\|_{L_p(I)}$. Letting $k$ tend to $+\infty$, we conclude that $W_p^p(\mu_0, \mu_T) = \|M_0 - M_T\|_{L_p(I)}$.

In fact, the $p^{\text{th}}$–Wasserstein space $(\mathcal{P}_p(\mathbb{R}), W_p)$ inherits several properties of the space $(L^p(I), \|\cdot\|_p)$. The fact that $L^p$–Wasserstein distance $W_p$ is a metric is a direct consequence of (4.17). We are now in the position to give a proof of the following:

**Theorem 4.46** For all $0 < p < +\infty$, the $L^p$–Wasserstein distance $W_p$ defined by the quantity (4.17) is, indeed, a norm over the $p^{\text{th}}$–Wasserstein space $(\mathcal{P}_p(\mathbb{R}), W_p)$.

**Proof.** We can assume $1 \leq p < +\infty$. Indeed, if $0 < p < 1$, we can replace $|\cdot|$ by the topologically equivalent $|\cdot|^p$. So, it is clear that $(\mathcal{P}_p(\mathbb{R}), W_p)$, $1 \leq p < +\infty$ is a normed space by the definition of $W_p$ (see Corollary 4.44 and Remark 4.45). \[\square\]

The proof above strongly uses the results about monotone nondecreasing maps in $L^p(I)$, which are transport maps from previous sections and is somehow specific to $\mathbb{R}$. In a general metric space, the tricky part would be to approximate arbitrary measures with measures such that the $L^p$–Wasserstein distance can be computed with maps instead of plans; absolutely continuous measures are not necessary to use, but we need nonatomic ones to apply Theorem 3.16 (see Chapter 5 of [104]). To give a more general result, Villani [115] provides a different
proof by using the so–called Gluing Lemma, which may happen to be more difficult for the reader who is not accustomed to disintegrations of measures (see Chapter 7 of [115] for more details).

Notice that, the Wasserstein distance $W_1$, which is the so–called by the Kantorovich–Rubinstein distance [115], is a special case of the duality theorem of Kantorovich and Rubinstein [80]. When $\mu_i \in \mathcal{P}(\mathbb{R}), i = 1, 2$ have bounded support, the dual representation of $W_1$ is

$$W_1(\mu_1, \mu_2) = \sup_{\xi \in \text{Lip}_1(\mathbb{R})} \int_{\mathbb{R}} \xi(x)(\mu_2(dx) - \mu_1(dx)),$$

where $\text{Lip}_1(\mathbb{R})$ is the set of Lipschitz functions on $\mathbb{R}$ whose Lipschitz constant not exceeding one.

**Proposition 4.47** For $p = 1$ one gets, thanks to Theorem 4.37, the following explicit expression of the $L^1$–Wasserstein distance

$$W_1(\mu_1, \mu_2) = \|M_1 - M_2\|_{L^1(I)} = \|F_1 - F_2\|_{L^1(\mathbb{R})},$$

where $F_1, F_2$ are the c.d.f. of $\mu_1, \mu_2$, respectively.

In fact, it follows that, the $L^p$–Wasserstein distance $W_p$ is a metrization of $C'(\mathbb{R})$ for any $1 \leq p < +\infty$, where $C'(\mathbb{R})$ is endowed with the $L^1$–Wasserstein distance. That is when all measures $\mu_k$ and $\mu$ are probabilities over $\mathbb{R}$, the weak convergence of the sequence $\{\mu_k\}_{k \in \mathbb{N}}$ to $\mu$ coincides with the convergence in the duality with $C_c(\mathbb{R})$. Note, however, that the metrics $W_{p_1}$ and $W_{p_2}$ are not equivalent unless $p_1 = p_2$. Since all $\gamma \in \Gamma(\mu_1, \mu_2)$ are probability measures, a simple application of Jensen inequality for $1 \leq p_1 \leq p_2$ implies $W_1 \leq W_{p_1} \leq W_{p_2}$. On the other hand, an opposite inequality also holds, since $1 \leq p_2 \leq p_1$ implies $W_{p_1}^{p_1}(\mu_1, \mu_2) \leq \text{diam}(\mathbb{R})^{p_1-p_2} W_{p_2}^{p_2}(\mu_1, \mu_2)$, where $\text{diam}(\mathbb{R}) = \sup\{|x_1 - x_2|: x_i \in \mathbb{R}, i = 1, 2\}$. Thus, all distances $W_p (p \geq 1)$ define the same topology on $\mathcal{P}(\mathbb{R})$.

**Lemma 4.48** If $p > 1$, then any bounded family in $\mathcal{P}_p(\mathbb{R})$ is uniformly $1 - 1/p$–Holder as a function form $[0, T]$ into $C'(\mathbb{R})$,

$$W_p(\mu_1, \mu_2) \leq C W_1^{1/p}(\mu_1, \mu_2) \quad \text{for some } C = \text{diam}(\mathbb{R})^{(p-1)/p}.$$
4.2.2 Topology of the $p^{th}$-Wasserstein space

As already mentioned, the $L^p$–Wasserstein distance $W_p$ is our main interest and our study during this thesis is only restricted on $\mathbb{R}$. However, some of the theorem presented next can be proved in a more general context. In this section, we want to analyze the convergence in the $p^{th}$–Wasserstein space on $\mathbb{R}$ and compare it to the notion of weak convergence. First of all, let us clarify that we often use the term “weak convergence”, when talking about probability measures, to denote the convergence in the duality with $C_b$, the space of bounded continuous functions (which is often referred to as narrow convergence and write $\mu_k \rightharpoonup \mu$) to say that $\mu_k$ converges in such a sense to $\mu$; that is, $\int \zeta\,d\mu_k \to \int \zeta\,d\mu$ as $k \to +\infty$ for any $\zeta \in C_b(\mathbb{R})$.

In the general case, once the weak convergence notation is understood, one can easily prove [104] that $\mu_k \rightharpoonup \mu$ in $\mathcal{P}_p(\mathbb{R})$ as $k \to +\infty$ if and only if $W_1(\mu_k, \mu) \to 0$. The proof of Theorem 4.49 below is given in [104], [116].

**Theorem 4.49** If $1 \leq p < +\infty$, then the $L^p$–Wasserstein distance $W_p$ metrizes the weak convergence in $\mathcal{P}_p(\mathbb{R})$. In other words, if $\{\mu_k\}_{k \in \mathbb{N}}$ is a sequence of measures in $\mathcal{P}_p(\mathbb{R})$ and $\mu$ is another measure in $\mathcal{P}(\mathbb{R})$, then $\mu_k \rightharpoonup \mu$ in $\mathcal{P}_p(\mathbb{R})$ if and only if $W_p(\mu_k, \mu) \to 0$.

Also, one can approximate $\mu_i$, $i = 1, 2$ by such discrete measures $\mu_i^k$, $i = 1, 2$, respectively in the $L^p$–Wasserstein distance $W_p$. The following corollary of Theorem 4.49, which illustrates the continuity of $L^p$–Wasserstein distance $W_p$, comes immediately from the triangle inequality:

**Corollary 4.50** If $1 \leq p < +\infty$, then the $L^p$–Wasserstein distance $W_p$ is continuous on $\mathcal{P}_p(\mathbb{R})$. More explicitly, if $\mu_1^k$ (resp. $\mu_2^k$) converges to $\mu_1$ (resp. $\mu_2$) weakly in $\mathcal{P}_p(\mathbb{R})$ as $k \to +\infty$, then $W_p(\mu_1^k, \mu_2^k) \to W_p(\mu_1, \mu_2)$.

What more can be said of the metric structure given by $L^p$–Wasserstein distance $W_p$? We finish this subsection with next statement that states that the $p^{th}$–Wasserstein space over $\mathbb{R}$, is a complete and separable space. In fact, $\mathcal{P}_p(\mathbb{R})$ is complete and locally compact under the weak $C_w$ topology if $p > 1$. That is, for any bounded sequence $\mu_k$ in $\mathcal{P}_p(\mathbb{R})$ we can extract a subsequence which converges in $C_w([0, T] \times \mathbb{R})$ to some $\mu \in \mathcal{P}_p(\mathbb{R})$ (see
for example proofs in [30], [116]). In addition, the proof of the theorem below holds for a more general class of spaces [17], [30], [115], [116]. In that case, \((\mathcal{P}_p(\mathcal{X}), W_p)\) is the so-called Polish spaces whenever a space \(\mathcal{X}\) is itself Polish. We give here an elementary proof that the \(L^p\)-Wasserstein distance \(W_p\), which play a basic role in optimal transportation issues, turn some spaces of probability measures \(\mathcal{P}_p(\mathbb{R})\) into separable complete metric spaces. Observe that a metric space is separable if it has a countable dense subset.

**Lemma 4.51** Any subset of a separable metric space is separable (with its relative topology).

**Proof.** Let \(\mathcal{X}\) be a separable metric space, with metric \(d\). Let \(\mathcal{C} = \{p_n : n \geq 1\}\) be a countable dense subset of \(\mathcal{X}\). Consider an infinite subset \(\mathcal{Y}\) of \(\mathcal{X}\). For all positive integers \(n\) and \(k\), if \(B(p_n, 1/k) \in \mathcal{Y}\) is nonempty, choose a point \(q(n,k)\) in this intersection. Let \(\mathcal{Z}\) be the collection of all the \(q(n,k)\)'s. Obviously \(\mathcal{Z}\) is a countable subset of \(\mathcal{Y}\). Furthermore, \(\mathcal{Z}\) is a dense subset of \(\mathcal{Y}\). For, let \(x\) be a point of \(\mathcal{Y}\), and \(k\) be a positive integer. Because \(\mathcal{C}\) is dense in \(\mathcal{X}\), there is a point \(p_k\) in \(\mathcal{C}\) with \(d(x, p_k) < 1/k\). Therefore, \(x\) is in \(B(p_k, 1/k)\), so \(B(p_k, 1/k)\) is nonempty and contains \(q(k,1/k)\), a point of \(\mathcal{Z}\). Then, by the triangle inequality, \(d(x,q(k,1/k)) < 2/k\), so \(\lim_{k \to +\infty} d(x,q(k,1/k)) = 0\). Accordingly \(\mathcal{Z}\) is dense in \(\mathcal{Y}\). \(\Box\)

**Theorem 4.52** The \(p^{th}\)-Wasserstein space \((\mathcal{P}_p(\mathbb{R}), W_p), 1 \leq p < +\infty\) endowed with the \(L^p\)-Wasserstein distance is a complete, separable metric space.

**Proof.** The proof follows easily in the one-dimensional case from the fact that the \(L^p\)-Wasserstein distance \(W_p\) is the \(L^p\)-norm (restricted to \(I\)) of monotone nondecreasing maps pushing forward \(\chi\) to \(\mu\) in \(\mathcal{P}_p(\mathbb{R})\) (see Corollary 4.44). The set \(\text{Mon}\) consisting of all monotone nondecreasing functions lying in \(L^p(I)\) is, in fact, separable, due to Lemma 4.51 (because \(L^p(I)\) is separable for all \(1 \leq p < +\infty\)). Moreover, the map \(\pi : L^p(I) \to \mathcal{P}_p(\mathbb{R})\) is surjective and \(1\)-Lipschitz with respect to \(W_p\) (this is proven in Proposition 7.29 of [116] for a locally compact length space). Furthermore, for \(M_1, M_2 \in L^p(I)\) we may choose \(\bar{M}_2 \in L^p(I)\) such that \(\pi(\bar{M}_2) = \pi(M_2)\) and \(W_p(\mu_1, \mu_2) = \|M_1 - \bar{M}_2\|_{L^p(I)}\). Therefore, the set \(\text{Mon}\) is a closed, convex subset of \(L^p(I)\) which is isometrically identical to \(\mathcal{P}_p(\mathbb{R})\). Thus, \(W_p\) defines a complete metric on \(\mathcal{P}_p(\mathbb{R})\). \(\Box\)
4.2.3 Convergence in sense of cumulative distribution functions

It is interesting to see how Theorem 4.49 can be proven directly in the one–dimensional case, in which case weak convergence can be expressed in terms of c.d.f. We first recall the following standard fact, whose proof can be found in almost any introductory textbook in probability theory (see for example Villani’s books [115], [116]).

**Theorem 4.53** Let \((\mu_k)_{k \in \mathbb{N}}\) be a sequence of probability measures on \(\mathbb{R}\), with respect to c.d.f. \(F_k\), and let \(\mu\) be a probability distribution with c.d.f. \(F\). Then, \(\mu_k\) converges to \(\mu\) in the weak sense if and only if \(F_k(x) \to F(x)\) as \(k \to +\infty\) for all points \(x \in \mathbb{R}\) at which \(F\) is continuous.

As consequence of this theorem, a sequence \((X_k)_{k \in \mathbb{N}}\) of real–valued random variables is said to converge weakly (or in distribution or in law), denoted by \(X_k \leadsto X\), to a random variable \(X\) as \(k \to +\infty\) if the sequence of pushforward measures \((X_k)_#\mathbb{P}\) converges weakly to \(X_#\mathbb{P}\) in the sense of weak convergence of measures on \(\mathbb{R}\).

Recall also that, \(W_1\) is computed with the usual distance on the real line (see Proposition 4.47). For simplicity let us assume that we deal with probability distributions on \(\mathbb{R}\) with common compact support. Taking into account the equivalence of all Monge–Kantorovich distances in this case, Theorem 4.53 translates into the following statement.

**Proposition 4.54** Let \((F_k)_{k \in \mathbb{N}}\) and \(F\) be nondecreasing, right–continuous functions on an interval \([0, 1] \subset \mathbb{R}\). Then, \(F_k \to F\) as \(k \to +\infty\) in the \(L^1\)–norm if and only if \(F_k(x) \to F(x)\) as \(k \to +\infty\) for all points \(x \in \mathbb{R}\) at which \(F\) is continuous.

In addition to the properties of the transform \(F \mapsto F^{-1}\) that we have mentioned in Subsection 4.1.1, there are other interesting ones such as the following one which is equivalent to the so-called Elementary Skorohod Theorem [107]. We provide a proof for completeness.

**Lemma 4.55** Let \((F_k)_{k \in \mathbb{N}}\) be a sequence of c.d.f., and assume that \(F_k \to F\) weakly, i.e. \(\lim_{k \to +\infty} F_k(x) = F(x)\) for any point \(x\) of continuity of \(F\). Then \(\lim_{k \to +\infty} F_k^{-1}(x) = F^{-1}(x)\), for any point \(t\) of continuity of \(F^{-1}\).
Proof. Let $y$ be a point of continuity of $F^{-1}$ and let $D \subset \mathbb{R}$ be the set of all continuity points of all $F_k$’s (which is dense on the real line). We first show that, given $\varepsilon > 0$, $F^{-1}(y) \leq F^{-1}(y) + \varepsilon$ for all $k$ large enough. By Lemma 4.3, this is equivalent to $F_k(F^{-1}(y) + \varepsilon) \geq y$. Choose $x \in D$ such that $F^{-1}(y) + \varepsilon > x > F^{-1}(y) + \varepsilon/2$. Then $F_k(x) \geq F(F^{-1}(y) + \varepsilon/2) - \delta$, for all $k$ large enough with any prescribed $\delta > 0$. Hence, it suffices to prove that $F(F^{-1}(y) + \varepsilon/2) \geq y + \delta$. But if $0 < \delta < 1 - y$, the latter is equivalent to $F^{-1}(y) + \varepsilon/2 \geq F^{-1}(y + \delta)$. An appropriate value of $\delta$ can be then chosen once $F^{-1}$ is continuous at the point $y$. By a similar argument, $F^{-1}(y) > F^{-1}(y) + \varepsilon$ for all $k$ large enough, concluding therefore the proof of the lemma. □

A direct consequence of Lemma 4.55 is Skorohod’s representation theorem which asserts the following:

**Theorem 4.56** If $(X_k)_{k \in \mathbb{N}}$ is a sequence of real–valued random variables, converging in distribution to a limit random variable $X$ (that is, the law $\mu_k$ of $X_k$ converges weakly to the law $\mu$ of $X$), then there exist random variables $(X'_k)_{k \in \mathbb{N}}$ and $X'$, such that for all $k$, $X'_k$ has the same law as $X_k$, and $X'$ has the same law as $X$, and $X'_k$ converges a.e. to $X'$.

**Sketch of Proof.** This theorem can be proven by elementary means: it suffices to set $X'_k = F_k^{-1}(U)$ and $X' = F^{-1}(U)$, where $U$ is uniformly distributed in $[0, 1]$, and to check the criterion given by Theorem 4.53. But with the help of the $L^1$–Wasserstein distance $W_1$, it is possible to give a more quantitative version of Skorohod’s theorem, in terms of $L^1$ norms: with the notation $E$ standing for expectation, under the additional assumption

$$\lim_{R \to +\infty} \limsup_{k \to +\infty} E[|X_k| 1_{|X_k| \geq R}] = 0$$

(which is satisfied for instance if $\sup_k E[|X_k|^p] < +\infty$ for some $p > 1$), we know from Theorem 4.53 that $W_1(\mu_k, \mu) \to 0$ as $k \to +\infty$, and we can write $\|X'_k - X'\|_{L^1(\mathbb{R})} = \|F_k^{-1} - F^{-1}\|_{L^1(0, 1)} = W_1(\mu_k, \mu)$. Hence we have shown the desired thesis. □

**Remark 4.57** The preceding Theorem 4.56 may be used, for example, to justify the identity (4.17) which is; in fact, a particular case of the identity (4.14). In the general case, one can approximate $\mu_i$, $i = 1, 2$ by such sequences measures $\mu_i^k$ in the metric $W_p$. If $F_i^k$, $F_i$, $i = 1, 2$
are the associated distribution functions, then necessarily $F_1^k \rightarrow F_1$ and $F_2^k \rightarrow F_2$ weakly. Hence, by Theorem 4.56 and Fatou’s lemma,

$$\|M_1 - M_2\|_{L^p(I)}^p \leq \liminf_{k \rightarrow +\infty} \|M_1 - M_2\|^p_{L^p(I)} = \liminf_{k \rightarrow +\infty} W_p^p(\mu_1^k, \mu_2^k) = W_p^p(\mu_1, \mu_2).$$

On the other hand, the joint distribution of $F_i^{-1}, i = 1, 2$ under the Lebesgue measure on $(0,1)$ has $\mu_i, i = 1, 2$ as marginals, respectively. Hence, by the definition of $L^p$–Wasserstein distance $W_p$ as Kantorovich transport distance, there is an opposite inequality

$$W_p^p(\mu_1, \mu_2) \leq \|M_1 - M_2\|^p_{L^p(I)}.$$
Chapter 5

Rearrangements and well-defined of the Lagrangian flows

In this Chapter, we begin by briefly recalling the results of the smooth and nonsmooth theory that we already mentioned in Chapters 1 and 2, here in Section 5.1 we motivate the extension in these contexts of an irregular (1D) vector field $v$, and we consider only the case of Borel velocity fields. In Subsections 5.1.1 and 5.1.2, we introduce the concept of Flow of a Borel map and the unidimensional continuity equation. In Subsection 5.1.3, we present some formal computations in order to motivate the notation of monotone rearrangement of the flows of a Borel maps, which will be central in the theory. Subsection 5.1.4 is devoted to the density of Lagrangian flows. In Section 5.2 we start to investigate the well–defined of the Lagrangian flows in Hilbert space $H^1(0, T; L^2(I))$, where $I$ is the unit open interval $(0, 1)$ on the real line. In Subsection 5.2.1 we recall some basic results [70] about optimal flows associated to absolutely continuous curves in $P_2(\mathbb{R})$ (the space of Borel probability measures with finite second–moments on $\mathbb{R}$). Subsection 5.2.2 is devoted to the notation of monotone rearrangement of the flows in the case of second–integrability. An illustrated example is also given to demonstrate the result of how this notation is applied with the $H^1$–regularity of the flow, i.e. each “Lagrangian” flow of a Borel map gives rise to a monotone solution in $H^1(0, T; L^2(I))$. In Subsection 5.2.3, we give examples of monotonicity and nonmonotonicity of flows where a very complicated counterexample to the nondecreasing monotonicity has been constructed. In Subsection 5.2.4, we give a first application of the theory of 1D Optimal
Transport that any Borel map defined on (in our case) $I$ can be monotonically rearranged over $I$. The second main application is the equivalence between paths in $H^1(0,T;L^2(I))$ consisting of monotone nondecreasing maps and curves in $AC^2(0,T;\mathcal{P}_2(\mathbb{R}))$, via the constraint “pushforward is prescribed”. We obtain a very strong statement of connection between the $H^1$–regularity of the optimal transport maps in 1D and the absolutely continuous curves with finite second–energy in the Wasserstein space $\mathcal{P}_2(\mathbb{R})$ without impose any assumptions on its associated velocity. The last section presents uniqueness of the Lagrangian flow associated to $AC^2(0,T;\mathcal{P}_2(\mathbb{R}))$ curves and some recent advances relative to it.

5.1 Lagrangian and Eulerian viewpoints in the 1D case

Motivated by the well–known results of Chapter 1 that for a given velocity vector field, which is merely a Borel function, we can give a definition to the Lagrangian flow of a Borel map. The reason for introduce this notion will become clear in Subsection 6.2.2 since it turns out to be the right one in the study of the ordinary differential equation with weakly differentiable solutions. Notice that the boundedness assumption on the vector field assumed in Chapter 1 is not essential: in this chapter we will also address the theory of Lagrangian flows relative to vector fields satisfying more general (Borel) regularities. It is, therefore, tempting to try here restricting our study for the Cauchy problem (1.3) was addressed in Chapter 1 on a specific “spatial” domain of $\mathbb{R}$ with generalizing to some initial data as a real–valued function. In this case, all observations in Chapter 1 about existence and uniqueness remain valid and the only different being that the identity function will be replaced by an arbitrary function as the initial point of the flow map $X$ on $[0,T] \times \mathbb{R}$.

5.1.1 Lagrangian flows of Borel velocity fields on the real line

Let us begin by investigating the following differential equation

$$\partial_t X(t,\cdot) = v(t,X(t,\cdot)), \quad \text{for every } t \in [0,T],$$

(5.1)
together, of coerce, with the prescribed initial datum

\[ X(0, \cdot) = X_0, \quad (5.2) \]

where \( I \) is the interval \((0,1)\) and \( X_0 : I \to \mathbb{R}, v : [0,T] \times \mathbb{R} \to \mathbb{R} \) are given (Borel) functions.

As observed earlier in Chapter 1, this equation describes the flow of a scalar field when the vector field \( v \) represents the fluid flow velocity at position \( X_0(x) \) (for every \( x \in I \)) at time \( t \). The fluid flow is described by a map \( X(t, \cdot) \) giving the position of the particle labeled \( X_0 \) at time \( t \). This equation is the basis for the Lagrangian description of fluid flow (where the trajectory of a particle is observed through time). Therefore, we take the opportunity to further motivate our study by pointing out that the proposed research throughout this thesis addresses the fundamentals of Lagrangian description of Fluid Dynamics. We are thus looking for characteristic (or integral) curves of the given velocity field \( v \), i.e., curves with the property that at each point the tangent vector coincides with the value of the given vector field at such point. We will call such Initial value (or Cauchy) problem \((5.1)-(5.2)\) above by a \emph{flow problem}.

Note that, Proposition 1.2 in Chapter 1 typically states that if we assume, for a moment, that the vector field \( v \) is sufficiently regular with respect to the time variable \( t \) and the spatial variable \( x \), then these assumptions guarantee that, the flow map is globally in time well–defined and that for every \( t \in [0,T] \), \( X(t, \cdot) \) is a smooth invertible and we denote by \( X^{-1}(t, \cdot) \) its inverse. In other words, for every \( x \in I \), there exists a unique global solution of the flow problem \((5.1)-(5.2)\); that is, this problem is a globally solvable for every \( x \in I \) by the only one “classical” flow \( X(t, \cdot) : I \to \mathbb{R} \) of the vector field \( v \) starting at \( X_0 \) when \( t = 0 \). Thus, the map \( X(t, x) \) solves the flow problem \((5.1)-(5.2)\) for every \( x \in I \). For mere existence the continuity of \( v \) is sufficient (see Proposition 1.2). Uniqueness can be also obtained under less stringent conditions such as \emph{onesided Lipschitz condition} or \emph{Osgood continuity}; according to Remark 1.4. Moreover, if \( v \in C^1_c([0,T] \times J) \), where \( J \subset \mathbb{R} \) is open interval, then the unique flow map \( X(t, \cdot) \) is a diffeomorphism of \( J [113], [114] \). More generally, assuming that \( v \in L^1_{loc}([0,T] \times \mathbb{R}) \) implying the uniqueness of solutions of the flow problem \((5.1)-(5.2)\) via
Theorem 1.5 as an equivalent to requiring that $X(\cdot, x)$ satisfy

$$X(t, x) = X_0(x) + \int_0^t v(s, X(s, x)) \, ds \quad \text{for } \mathcal{L}^1\text{-a.e } x \in I \text{ and every } t \in [0, T]. \quad (5.3)$$

It is convenient to introduce now the definition of Flow of a Borel map, which is Lagrangian flow restricted to the one–dimensional case. If $\mathcal{L}^1$ is the Lebesgue measure on $\mathbb{R}$, then we take for $I$ the unit open interval $(0, 1)$ of the real $\mathbb{R}$ to ensure that the pushforward (or image) of measure defined on $I$ is a probability measure; that is $X_\#\mathcal{L}^1|_I \in \mathcal{P}(\mathbb{R})$, but everything we achieve in this thesis can be trivially extended to any bounded interval.

**Definition 5.1 (Flow of a Borel map)** Let $v : [0, T] \times \mathbb{R} \to \mathbb{R}$ and $X_0 : I \to \mathbb{R}$ be Borel functions. We say that $X : [0, T] \times I \to \mathbb{R}$ is a flow map for $v$ if

(i) The map $[0, T] \ni t \to X(t, x)$ is absolutely continuous for $\mathcal{L}^1\text{-a.e. } x \in I$;

(ii) $\partial_t X(\cdot, x) = v(\cdot, X(\cdot, x))$ for $\mathcal{L}^1\text{-a.e. } x \in I$.

Furthermore, we say that $X : [0, T] \times I \to \mathbb{R}$ is a flow map for $v$ starting at $X_0$ if, beside (i), (ii), the following is satisfied:

(iii) $X(0, x) = X_0(x)$ for $\mathcal{L}^1\text{-a.e. } x \in I$.

Notice that, the map $X$ need not be differentiable at all times, but at $\mathcal{L}^1\text{-a.e. } t \in [0, T]$ because, for each $\mathcal{L}^1\text{-a.e. } x \in (0, 1)$, it is absolutely continuous function in time, which has a time derivative for $\mathcal{L}^1\text{-a.e.}$. The main observation in this subsection is that, (ii) and (iii) of the definition above turn out to be equivalent to requiring that $X(\cdot, x)$ satisfy the integral equation (5.3) above. Also, the absolutely continuity (i), according to (ii), is equivalent to $[0, T] \ni t \to v(t, X(t, x)) \in L^1(0, T)$ for $\mathcal{L}^1\text{-a.e. } x \in I$.

### 5.1.2 The unidimensional continuity equation

We now present our approach to the existences and/or uniqueness theory of Lagrangian flows, based on the concept of (mass) density transported by the flow and particularly useful to deal with a Borel velocity vector fields (recall Definition 5.1 above and take a different
look to the approach in Chapter 1 for the PDE theory in this framework). The presentation here is very different to the one by Ambrosio [6], DiPerna–Lions [56], and we also refer to De Lellis [50], [51] for a complete exposition. This approach works naturally under an absolute continuity assumption to the probability measures on the real line, which is a bit weaker than the one assumed in the standard definition of regular Lagrangian flow (see Definition 2.8(ii)): we will only need that

$$X_t \# \mathcal{L}^1 \ll C \mathcal{L}^1 \quad \text{for every } \mathcal{L}^1\text{-a.e. } t \in (0, T).$$

This condition is equivalent to saying that: for \( \mathcal{L}^1\text{-a.e. } t \in (0, T) \), we have \( \mathcal{L}^1(\{x \in \mathbb{R} : (x, X_t(x)) \in B\}) = 0 \) for every Borel set \( B \subset \mathbb{R} \) with \( \mathcal{L}^1(B \setminus \mathbb{R}) = 0 \). Let us further restrict our attention to measures \( (0, T) \ni t \mapsto \mu_t \) of the form \( \mu_t(B) = \chi(\{x \in I : (x, X_t(x)) \in B\}) \) with \( \chi := \mathcal{L}^1|_I \in \mathcal{P}_p(\mathbb{R}) \) and \( X_t : I \to \mathbb{R} \) in \( L^p(I) \) with \( 1 \leq p < +\infty \). More precisely, if \( \mathcal{L}^2 \) is the Lebesgue measure on \( \mathbb{R}^2 \) and \( X \) is a Lagrangian flow associated to a Borel vector field \( v \), we define

$$\mu = (\text{Id}, X) \# \mathcal{L}^2 \ll ((0, T) \times I),$$

where \( \text{Id} \) stands for the identity map on \( (0, T) \), i.e. \( \mu \) is the pushforward of the Lebesgue measure restricted on \( (0, T) \times I \) via the map \( (t, x) \mapsto (t, X_t(x)) \). Such pushforward is simply defined by the property \( \int_0^T \int_{\mathbb{R}} \xi(t, x) \, d\mu_t(x) \, dt = \int_0^T \int_{X_t} \xi(t, X_t(x)) \, dx \, dt \) valid for every \( \xi \in C_c(\mathbb{R}^2) \) (see Definition 3.2).

As already observed in Chapter 1, in the case of smooth vector field, if \( \mu_t := X_t \# \mathcal{L}^1|_I \) (i.e. \( \mu_t(B) = \chi(X_t^{-1}(B)) \)) for all Borel sets \( B \subset \mathbb{R} \), where \( \chi \) is the one–dimensional Lebesgue measure restricted on \( I \), then we get the Eulerian description of the distribution \( X_0 \# \chi \) as it is transported by the velocity \( v \): this means \( (\mu, v) \) solves the unidimensional continuity equation

$$\partial_t \mu + \partial_y(\mu v) = 0 \quad \text{in } \mathcal{D}'((0, T) \times \mathbb{R})$$

from Fluid Mechanics. It is worth mentioning that the flow problem (5.1)–(5.2) and (5.5)
are closely connected to the unidimensional transport equation

\[ \partial_t F + v \partial_y F = 0 \quad \text{in} \quad (0, T) \times \mathbb{R}, \tag{5.6} \]

where \( F \) is the cumulative distribution function, even though we shall not pursue this in our present study.

**Remark 5.2** Observe that the c.d.f. \( F_t \) formally satisfies the transport equation (5.6) (see Proposition 7.3 below). We will see in Remark 7.4 that, the quantity \( g(t, x) := F(t, X(t, x)) \) of measure–preserving map, which pushes \( \chi \) forward to itself, is constant with respect to time. Hence, \( g(t, \cdot) \equiv 0 \), so \( g \) does not vary in time. This means that (5.6) can be uniquely solved thanks to the theory of characteristics (see Subsection 1.3.3): if we couple the transport equation (5.6) with an initial data \( F(0, x) = F_0(x) \), then the unique solution \( F_t(x) \) is the transport of the initial data \( F_0(x) \) along solutions of the flow problem (5.1)–(5.2), more precisely \( F_t(x) = F_0(X_0(t, x)) \). Thus, the a.e. unique \( \chi \)-preserving map \( g \) also satisfying the differential equations (3.30) and (3.31).

Now, let us restrict the results in Chapter 1 for the existence and uniqueness of solution for the differential equation (1.3) on the interval \((0, 1)\). Assume now, we are given a probability measure \( \mu_0 := \chi \) on \((0, 1)\). If \( v \) further satisfies the no–flux boundary condition: \( \partial_x v_t(x) = 0 \) for all \((t, x) \in (0, T) \times \mathbb{R} \), then we can prove the following proposition that we state for simplicity in the case of Lipschitz and bounded vector fields.

**Proposition 5.3** Suppose that \( I := (0, 1) \). Suppose that \( v_t : \mathbb{R} \rightarrow \mathbb{R} \) is Lipschitz continuous in \( x \), uniformly in \( t \), and uniformly bounded, and consider its flow \( X_t \). Suppose that for every \( x \in \mathbb{R} \) and every \( t \in [0, T] \), we have \( X_t \in \mathbb{R} \). Then, for the probability \( \chi \in \mathcal{P}(\mathbb{R}) \), the measures \( \mu_t := X_t \# \chi \) solve (distributionally) the identity (5.5) with initial datum \( \mu_0 = \chi \). Moreover, every solution of the same equation with \( \mu_t \ll \chi \) for every \( t \) is necessarily obtained as \( \mu_t = X_t \# \chi \). In particular, the continuity equation admits a unique solution.

**Proof.** First let us check the validity of the equation (5.5) when \( \mu_t \) is obtained from such a flow through the problem (5.1)–(5.2). Let us suppose that \( \text{spt}(\mu_t) \subset \mathbb{R} \) (which is satisfied
since \( \mu_0 := \chi \) is concentrated on the interval \( I := (0, 1) \) and \( v \) satisfies the no–flux boundary condition). We will check that we have a weak solution. In order to check (1.11), by the \( C^1 \)– regularity we only need to compute the pointwise derivative. Since the flow satisfies the problem (5.1)–(5.2) for every \( t \) and \( x \) we can deduce, for every a test function \( \xi \in C^1(\mathbb{R}) \) such that both \( \xi \) and \( \xi' \) are bounded
\[
\frac{d}{dt} \langle \xi, \mu_t \rangle = \frac{d}{dt} \int_I \xi(X_t(x)) \, dx = \int_I \xi'(X_t(x)) \dot{X}_t(x) \, dx = \int_I \xi'(X_t(x)) v_t(X_t(x)) \, dx = \langle \xi', \mu_t \, v_t \rangle,
\]
which characterizes weak solutions, i.e it proves that \( \partial_t \mu_t = -\partial_x (\mu_t \, v_t) \), in the weak sense.

Now, in order to prove the second part of the statement, we first observe that
\[
\int_0^T \int_\mathbb{R} \left( \partial_t \xi(t,x) + v_t(x) \partial_x \xi(t,x) \right) \, d\mu_t(x) \, dt = 0, \tag{5.7}
\]
is also valid for all Lipschitz compactly supported test functions \( \xi \in C^1_c((0, T) \times \mathbb{R}) \), whenever \( \mu_t \ll \chi \) for every \( t \). Indeed, if we fix a Lipschitz test function \( \xi \) and we smooth it by convolution with a compactly supported kernel, we have a sequence \( \xi^\varepsilon \in C^\infty_c \) such that \( \partial_{t,x} \xi^\varepsilon \to \partial_{t,x} \xi \) a.e. and we can apply dominated convergence since \( |\partial_{t,x} \xi^\varepsilon| \leq \text{Lip}(\xi) \) (we need \( \mu_t \) to be absolutely continuous because we only have Lebesgue–a.e. convergence).

Then we take a test function \( \zeta \in C^1_c(\mathbb{R}) \) and we define \( \xi(t,x) = \zeta(X_t^{-1}(x)) \). If we can prove that \( t \mapsto \langle \xi, \mu_t \rangle := \int_\mathbb{R} \xi(t,x) \, d\mu_t(x) \) is constant, then we have proven \( \mu_t = X_t \# \chi \). The function \( \xi \) is Lipschitz continuous, because the flow \( X_t \) is bi–Lipschitz; yet, it is not compactly supported in time (it is compactly supported in space since \( \text{spt}(\xi) \) is compact, and if we set \( L := \sup_{t,x} |v_t(x)| \), we see that \( \xi(t,x) \) vanishes on all points \( x \) which are at distance larger than \( tL \) from \( \text{spt}(\xi) \)). Thus, we multiply the continuity equation (5.5) with a cut–off function \( \phi(t) \), with \( \phi \in C^1_c([0, T]) \). We have
\[
\partial_t (\phi \, \xi) + v_t \partial_x (\phi \, \xi) = \phi'(t) \xi(t,x) + \phi(t) \left( \partial_t \xi(t,x) + v_t(x) \partial_x \xi(t,x) \right).
\]
We can prove that, by definition of \( \xi \), the term \( \partial_t \xi(t,x) + v_t(x) \partial_x \xi(t,x) \) vanishes a.e. (this corresponds to saying that \( \xi \) is a solution of the transport equation (5.6); see also observations in Section 1.3.3). This is true since we have \( \xi(t, X_t(x)) = \zeta(x) \), and differentiating it with
respect to \( t \) (which is possible for a.e. \((t, x)\)), we get \( \partial_t \xi(t, X_t(x)) + v_t(X_t(x)) \partial_x \xi(t, X_t(x)) = 0 \), which means that \( \partial_t \xi + v_t \partial_x \xi \) vanishes everywhere, as \( X_t \) is surjective. Hence, from the definition (5.7) of distributional solution, we have

\[
\int_0^T \phi'(t) \, dt \int_{\mathbb{R}} \xi(t, x) \, d\mu_t(x) = \int_0^T dt \int_{\mathbb{R}} \left( \partial_t(\phi \xi) + v_t(\partial_x(\phi \xi)) \right) \, d\mu_t(x) = 0.
\]

The test function \( \phi \in C^1_c([0, T]) \) being arbitrary, we get that \( t \mapsto \langle \xi, \mu_t \rangle \) is constant. \( \square \)

Note that, our endeavor from now on will be to study the flow problem (5.1)–(5.2) under the constraint that \( X_t \# \chi = \mu_t \) be given for all \( t \in [0, T] \) in the form of an absolutely continuous curve of probability measures [17].

### 5.1.3 A strategy to obtain uniqueness of the Lagrangian flows

As already observed in Chapter 1, if \( v = v(t, y) \) is sufficiently smooth, one can formally take the derivative of both sides of the flow equation (5.1) with respect to \( x \) to verify that the spatial derivative \( u(t, x) := \partial_x X(t, x) \) satisfies

\[
\partial_t u(t, x) = \partial_y v(t, X(t, x)) \, u(t, x), \quad u(0, x) = X'_0(x).
\] (5.8)

By letting \( g(t, x) := \partial_y v(t, X(t, x)) \) and assuming \( X \) is known, one can deal with \( \dot{u}(t, x) = g(t, x)u(t, x) \) as if it were a scalar linear equation in \( u \), with \( x \) being just a parameter. The solution procedure shows that

\[
u(t, x) = X'_0(x) \exp\left( \int_0^t g(s, x) \, ds \right)
\] (5.9)

is the unique solution to (5.8). Whereas we have not really found an explicit solution (because \( g \) depends on \( X \)), we can conclude from (5.9) that \( X'_0(x) > 0 \) implies \( \partial_x X(t, x) > 0 \) for all \((t, x) \in (0, T) \times I\), which shows that for every \( t \in (0, T) \), the function \( x \mapsto X(t, x) \) is strictly increasing, with positive slope everywhere. This, when referring to (5.5), is in agreement with the intuition that “nice” velocities \( v \) preserve in time the order of the positions of particles on the real line, or, equivalently, the characteristics do not cross. If \( v \) is not smooth enough to justify (5.8), then the order need not be preserved through time. This has fundamental
implications in scalar Conservation Laws, where the crossing of characteristics is responsible for the formation of shocks.

Our goal in this work is to study conditions under which $x \mapsto X(t, x)$ stays monotone nondecreasing for all times $t \in [0, T]$ if $X_0$ is itself monotone nondecreasing. The meaning of “conditions” is quite vague at this point; however, they shall not be imposed on the velocity $v$, but rather on the Eulerian flow (5.5) whose Lagrangian characterization is provided by the problem (5.1)–(5.2). Under the minimal assumptions we shall impose on $\mu$, the velocity $v$ will be uniquely determined by $\mu$ from (5.5) (in some precise sense).

5.1.4 The density of Lagrangian flows on the real line

We already have seen in the case of a smooth vector field $v$, the density $\rho$ of the image measure $X_0#\rho_0$ is obtained from the initial density $\rho_0$ through a simple change of variable involving the Jacobian factor (see Proposition 1.11). A standard approximation argument settles the case of nonsmooth vector fields. In general, however, an absolute continuity assumption for the density does not need to have any regularity on a Borel velocity vector fields. We observe that, thanks to Proposition 1.8, we can always assume $\rho \in C([0, T]; L^{\infty}_w(\mathbb{R}))$. This means that we recall that, up to a redefinition in a negligible set of times, every solution to (5.5) belongs to $C([0, T]; L^{\infty}_w(\mathbb{R}))$ (see Corollary 1.19).

**Remark 5.4** If $X$ is a Lagrangian flow of a Borel velocity map $v$ in the sense of Definition 5.1, we define $\rho_t = X_t#\chi$, i.e $\rho_t$ is the pushforward of the Lebesgue measure on $[0, T] \times \mathbb{R}$ via the map $(t, x) \rightarrow (t, X(t, x))$. This means that, thanks to Remark 4.28, we have the Monge–Ampère equation $\rho_t(X_t(x)) \partial_x X_t(x) = 1$ for every $x \in I$: if the vector field $v$ is smooth, take for $\rho$ a reciprocal of the spatial derivative of the flow $X$ generated by $v$; that is, $\rho(t, x) = 1/\partial_x X_0(t, x) = 1/\partial_x X_t(X_t^{-1}(x))$ for every $x \in I$, which is bounded away from zero, in light of (5.9) above, since we can compute explicitly $\rho(t, x) = \exp \left( - \int_0^t \partial_x v_\tau (X_\tau(t, x)) \, d\tau \right) / X_0'(x)$, where $X_t(s, x)$ satisfies $\partial_t X_t(s, x) = v_t(X_t(s, x))$ and $X_t(t, x) = X_0(x)$. In general, however, an absolute continuity assumption for the density does not need to have any regularity on the Borel velocity map $v$. 
The previous remark justifies the following definition, which will be important in the discussion of next chapters.

**Definition 5.5** The function $\rho$ defined in the above remark is called the density of the Lagrangian flow $X$ of a Borel velocity map $v$ in the sense of Definition 5.1.

Now, in the spirit of Proposition 5.3, the notion of measure-valued solutions $(0, t) \ni t \mapsto \mu_t \in \mathcal{P}(\mathbb{R})$ of the continuity equation (5.5) can be extended to the case of density $\rho_t$ of the Lagrangian flow $X$ of Borel velocity maps $v$ in the sense of Definition 5.1; we only need to understand the density in the above definition as it is a solution of (5.5). Moreover, the initial condition $\mu(0, \cdot) = \mu_0$ can be prescribed in the following sense: $(\rho \mathcal{L}^1)(0, \cdot) = \rho_0 \mathcal{L}^1$. Observe that the “compressibility” condition (5.4) means that $\mu_t$ is the absolute continuity with respect to the Lebesgue measure, and hence gives the existence of a function $\rho \in L^1_{\text{loc}}((0, T) \times \mathbb{R})$ such that $\mu = \rho \mathcal{L}^2$ (due to Radon–Nikodym theorem). The following proposition establishes a link between this notion and the Eulerian side of the problem, namely with the continuity equation relative to the Borel velocity map $v$.

**Proposition 5.6** Let $X$ be a Lagrangian flow of Borel velocity map $v \in L^\infty((0, T) \times \mathbb{R})$ and let $\rho_0 \in L^\infty(\mathbb{R})$. Define $\mu = (\text{Id}, X)_#(\rho_0 \mathcal{L}^2)$. Then, there exists $\rho \in L^1_{\text{loc}}((0, T) \times \mathbb{R})$ such that $\mu = \rho \mathcal{L}^2$ and the identity $\partial_t \rho + \partial_x (v \rho) = 0$ holds in the sense of distributions in $(0, T) \times \mathbb{R}$ under $\rho(0, \cdot) = \rho_0$.

**Remark 5.7** The probability density of the Lagrangian flow $X$ of a Borel velocity map $v$ corresponds to the case $\rho_0 = \chi$ (where $\chi := \mathcal{L}^1|_I$ and $I := (0, 1)$) in the previous proposition: this means that the probability density $\rho$ satisfies $\partial_t \rho + \partial_x (v \rho) = 0$ under $\rho(0, \cdot) = \chi$. In particular, we deduce that, when a vector field $v$ is smooth enough to satisfy our little calculation above, the probability density of a flow $X$ relative to a Borel velocity $v$ is the density generated by $v$ in the sense of Proposition 5.6; that is, the unique solution $\rho \in L^\infty((0, T) \times \mathbb{R})$ of $\partial_t \rho + \partial_x (v \rho) = 0$ under $\rho(0, \cdot) = \chi$.

Note that, in general, the notion of weak solutions of (5.5) for a Borel vector field $v$ depends on its density $\rho$. However, it was proved in [50] that if $v$ has renormalization
property (see Definition 1.23) with some fixed density $\rho$, then the notion of weak solution of \((5.5)\) is independent of the choice of $\rho$. For Borel vector fields with the absolute continuity of probability densities it is possible to develop a well–defined theory of the Lagrangian flows arguing as in the proof of Theorem 5.20. This is precisely the link between the Lagrangian and the Eulerian problems: the proof of the theorem, relative to the well–defined for the Lagrangian flow $X$ of a Borel velocity map $v$ in the sense of Definition 5.1, is strongly based on the well–posedness for the continuity equation in Proposition 5.6. Indeed, the result of Theorem 5.20 implies uniqueness for both of the flow equation \((5.1)–(5.2)\) and the continuity equation \((5.5)\) for a Borel velocity map $v$ with the absolute continuity of densities $\rho$. Moreover, the proof is essentially based on a monotone rearrangement procedure, which however requires some care in the next section, and for this, we refer the reader to take a look at some general techniques that could be exploited to complete the proof.

5.2 Lagrangian flows in the $H^1(0, T; L^2(I))$ space

The most general problem \((5.1)–(5.2)\) discussed here in this section assumes only Borel regularity on $v$, and study solutions $X$ only in the Hilbert space $H^1(0, T; L^2(I))$ (we call it $H^1$ in time). Maps in $H^1(0, T; L^2(I))$ posses a “functional” time–derivative in $L^2((0, T) \times I)$, which, in fact, coincides on a subset of $(0, T) \times I$ of full Lebesgue measure with a pointwise “pseudo”–derivative [70]. The flow equation \((5.1)\) is to be understood in that context, i.e. it should be satisfied everywhere in the set where the functional derivative $\dot{X}$ coincides with the pointwise pseudo–derivative. This definition turns out to be equivalent to requiring that $X(\cdot, x)$ satisfy the identity \((5.3)\).

5.2.1 Optimal flows associated to $AC^2(0, T, \mathcal{P}_2(\mathbb{R}))$ curves

Let us recall some basic facts from the theory of $L^2$–absolutely continuous curves in $\mathcal{P}_2(\mathbb{R})$ (the space of Borel probability measures on $\mathbb{R}$ with finite second–order moments). We shall be quite sketchy within the one–dimensional setting $d = 1$, for further details in the higher dimensional setting we recommend the comprehensive reference [17]. Suppose $\mu_i \in \mathcal{P}_2(\mathbb{R}), i = 1, 2$ and let $M_i : (0, 1) =: I \to \mathbb{R}$ be the unique a.e. monotone nonde-
creasing maps such that $M_i\chi = \mu_i$, respectively, where $\chi$ is the one-dimensional Lebesgue measure restricted to $I$. Then, $W_2(\mu_1, \mu_2) = \|M_1 - M_2\|_{L^2(I)}$. Let us endow $\mathcal{P}_2(\mathbb{R})$ with the quadratic Wasserstein metric $W_2$. Thus, $(\mathcal{P}_2(\mathbb{R}), W_2)$ becomes a Polish space on which we define absolutely continuous curves by saying that $[0, T] \ni t \mapsto \mu_t := X_t\chi \in \mathcal{P}_2(\mathbb{R})$ lies in $AC^2(0, T; \mathcal{P}_2(\mathbb{R}))$ provided that there exists $\beta \in L^2(0, T)$ such that $W_2(\mu_t, \mu_{t+h}) \leq \int_{t}^{t+h} \beta(s) \, ds$ for all $0 < t < t + h < T$ (see Definition 6.2 below). The metric derivative of such a curve is defined as

$$|\mu'|(t) = \lim_{s \to t} \frac{W_2(\mu_s, \mu_t)}{|s - t|} \quad \text{for } L^1\text{-a.e. } t \in (0, T).$$

There exists a unique [17] Borel velocity field $v : (0, T) \times \mathbb{R} \to \mathbb{R}$ such that $(\mu, v)$ satisfies (5.5) in the sense of distributions and $\|v\|_{L^2(\mu_t)} = |\mu'|(t)$ for a.e. $t \in (0, T)$. We call $v$ a velocity associated to the path $\mu$. Note that, due to Lemma 4.4 of [92] (see also Theorem 6.5 below), the minimality of the $L^2(\mu_t)$-norm as a selection principle is unnecessary here:

If a Borel map $w : (0, T) \times \mathbb{R} \to \mathbb{R}$ is another velocity field associated to the path $\mu$, then for $L^1$–a.e. $t \in (0, T)$ we have $\|v\|_{L^2(\mu_t)} = \|w\|_{L^2(\mu_t)}$.

The following statement appears in [70]. The main purpose of recalling this statement is to present a similar result in the case of higher integrability (see Theorem 6.33).

**Proposition 5.8** Suppose $\mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R}))$. Let $v$ be the velocity associated to $\mu$ and $M_t : I \to \mathbb{R}$ be monotone nondecreasing map such that $M_t\chi = \mu_t$. For each $t$, modifying $M_t$ on a countable subset of $I$ if necessary, we may assume without loss of generality that $M_t$ is left continuous. We have that

$$v_t(M_t(x)) = \dot{M}_t(x) \quad \text{for } L^2\text{-a.e. } (t, x) \in (0, T) \times I. \quad (5.10)$$

**Sketch of the proof.** The proof [70] is based on the observation that if $\mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R}))$ and $M_t : I \to \mathbb{R}$ are as in the statement of the proposition, then

For Lebesgue almost all $(t, x) \in (0, T) \times I$, $M_t(x) = M_t(y)$ implies $\dot{M}_t(x) = \dot{M}_t(y) \quad (5.11)$

(in case both derivatives exist pointwise in some sense, later to be specified). Furthermore,
the proof uses the fact that if $\mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R}))$ and $M_t : I \to \mathbb{R}$ is a monotone nondecreasing map such that $M_{t\#\chi} = \mu_t$ for all $t \in [0, T]$, then $M \in H^1(0, T; L^2(I))$ and the metric derivative $|M'(t)|$ exists at $t \in (0, T)$ if and only if the metric derivative $|\mu'(t)|$ exists at $t$; in that case $|M'(t)| = |\mu'(t)|$. Note also that (5.11) is a necessary condition for (5.10) to hold.

$\Box$

5.2.2 Monotone rearrangement of the optimal flows

It has been recently proved in [70] that, under some conditions, as soon as the flow problem (5.1)–(5.2) has a solution $X$, then one automatically gets a monotone solution obtained by setting $M_t := Mon(X_t)$ for all $t \in [0, T]$, where $Mon(S)$ stands for the monotone rearrangement of the map $S$. It is basically, this define that for each time $t \in [0, T]$, the map $M_t$ is a monotone rearrangement of the map $X_t$. This means that $M_t$ satisfies the flow problem (5.1)–(5.2) in the sense of Definition 5.1. Note that, it is not clear what do we mean by this meaning, since $M_t$ is just a monotone rearrangement of a given fixed map $X_t$ that we know by Proposition 5.8 above it might be a solution of the the flow problem (5.1)–(5.2) in the sense of Definition 5.1 (see the flow–map formula (5.10) above). Indeed, thanks to Definition 3.26 and Theorem 3.32, we know the meaning of the rearrangement of monotone solutions $X_t$ for the flow, which we call it $M_t$, but we still need a family of curves of probabilities $\mu := \{\mu_t\}_{t \in [0, T]}$ to take optimal maps $X_{t\#\chi} = \mu_t$ that pushing forward a measure $\chi$ into a family of curves of probabilities $\mu_t$, so $M := \{M_t\}_{t \in [0, T]}$ is a time–dependent family of monotone rearrangements of those optimal maps $X_t$ for all $t \in [0, T]$. This is a very ambiguous right now but it will be more precisely and more clearly later (see Theorems 6.33 and 6.35 below). Starting with Example 5.10 below we try to illustrate all these concepts and find relevant facts. The result that was proved reads:

**Theorem 5.9** Suppose $X_0 : I \to \mathbb{R}$ is monotone nondecreasing map. If $X$ belongs to $H^1(0, T; L^2(I))$ and solves the flow problem (5.1)–(5.2), then so does $M$.

**Sketch of the proof.** The proof is based on tools developed in [70], [92]. A crucial observation is that whenever $X \in H^1(0, T; L^2(I))$, the path $[0, T] \ni t \to \mu_t := X_{t\#\chi} \in \mathcal{P}_2(\mathbb{R})$ (denotes the set of all Borel probability measures with finite second moment) belongs to $AC^2(0, T; \mathcal{P}_2(\mathbb{R}))$
and \(v\) is its associated velocity field (see [17]) such that \((0, T) \ni t \rightarrow \|v_t\|_{L^2(\mu_t)} \in L^2(0, T)\). But the flow problem (5.1)–(5.2) readily implies that \((\mu, v)\) solves the continuity equation (5.5) from Fluid Mechanics in the scene of distributions. We have seen ([17] and [92]) that given \(\mu, v\) is unique with this property provided that \(v_t \in L^2(\mu_t)\) for a.e. \(t \in (0, T)\) (which holds, due to the flow problem (5.1)–(5.2)). Thus, \(v\) coincides with the velocity of minimal norm [17] along \(\mu\). Finally, by Proposition 5.8, we conclude the proof.

Notwithstanding the imposed \(H^1\)–regularity, Theorem 5.9 basically states that, if \(X_0\) is nondecreasing, then as soon as the flow problem (5.1)–(5.2) admits a solution \(X \in H^1(0, T; L^2(I))\), it automatically admits the solution \(M \in H^1(0, T; L^2(I))\) consisting of the monotone rearrangements of the maps \(X_t\) (which, in this context, we take it to mean that \(M_t\) is monotone nondecreasing for all \(t \in [0, T]\)). We have seen in Subsection 5.1.3 that any solution \(X\) of the flow problem (5.1)–(5.2) will be nondecreasing provided that \(X\) and \(v\) are regular enough to justify (5.8). It is, however, this regularity that might not be present there, which will enable solutions that are not nondecreasing in \(x\) (see Example 5.10 below that will illustrate this monotonicity property). Examples 5.13 and 5.14 however, illustrate that there are even irregular \(v\) for which only nondecreasing solutions exist. Thus, it comes as a very natural question to investigate when the flow problem (5.1)–(5.2) has only monotone solutions. The subtlety of the matter will become very clear, as we shall see that a careful formulation is in order.

A study of the regularity of the monotone rearrangements of maps belonging to a time-continuous family of maps was performed by Loeper in [85]. It is proved that the distributional time derivative is a signed measure. This is very weak by comparison to what we achieve here, but it covers a much more general case: arbitrary spatial dimension. Moreover, the family of maps is not necessarily the flow of a map.

**Example 5.10** Let us consider the (Borel) function \(v : [0, 2] \times \mathbb{R} \to \mathbb{R}\) given by

\[
v(t, y) = \frac{y}{t - 1} \quad \text{if } t \in [0, 2] \setminus \{1\}, \quad v(1, \cdot) \equiv 0.
\]

Clearly, \(v\) is analytic everywhere except on the fiber \(\{1\} \times \mathbb{R}\). It is easy to see that \(X(t, x) := (1 - t)x\) satisfies \(X \in H^1(0, T; L^2(I))\) (it is, in fact, analytic and bounded in \((0, 2) \times I\) and...
is a solution for the flow problem (5.1)–(5.2) with \(X_0(x) = x\) for this particular \(v\). Note that \(X_t\) is nondecreasing only if \(t \in [0, 1]\), and strictly decreasing otherwise. But, in agreement with Theorem 5.9, by setting

\[
M(t, x) = (1 - t)x \quad \text{if } t \in [0, 1] \quad \text{and} \quad M(t, x) = (1 - t)(1 - x) \quad \text{if } t \in (1, 2],
\]

we observe that \(M \in H^1(0, T; L^2(I))\) is also a solution with \(M_0(x) = x\) for all \(x \in I\). The maps \(M_t\) are the monotone rearrangements of the maps \(X_t\). The first branch of the piecewise function \(M\), for all \(t \in [0, 1]\) and all \(x \in I\), is exactly the same as the map \(X\) given above, but the second one when \(t \in (1, 2]\) is different. It is not difficult to check, by Definition 3.26, this map \(M(t, x) = (1 - t)(1 - x)\) for \(t \in (1, 2]\) and for \(x \in I\), is a monotone rearrangement of the map \(X(t, x) = (1 - t)x\). In fact, this map is increasing since, when we take the derivative with respect to \(x\), we get \(\partial_x M(t, x) = t - 1\), which is positive in the interval \((1, 2]\).

Let us fix a time horizon \(0 \leq t \leq 2\) to be; for example, \(t = 2\) and plug it in the equation \(X(t, x) = (1 - t)x\), we get \(X(2, x) = -x\) for all \(x \in I\), which is obviously decreasing, but when we plug it in \(M(t, x) = (1 - t)(1 - x)\), we get \(M(2, x) = x - 1\), which is increasing for all \(x \in I\). We know that, thanks to by Remark 3.27, the definition of monotone rearrangement, \(\chi(M_t^{-1}(B)) = \chi(X_t^{-1}(B))\) for some Borel set \(B \subset \mathbb{R}\), is not easy to check, but this turn out to be equivalent to the fact of integral \(\int_I \xi(M_t(x)) \, dx = \int_I \xi(X_t(x)) \, dx\) for any test function \(\xi\). We change variables in the left hand side of this integral equality to obtain

\[
\int_0^1 \xi((1 - t)(1 - x)) \, dx = \int_1^0 \xi((1 - t)y)(-dy) = \int_0^1 \xi((1 - t)y) \, dy = \int_0^1 \xi(X_t(x)) \, dx.
\]

Thus, this is an example where the flow problem (5.1)–(5.2) has other solutions beside monotone nondecreasing ones. It also illustrates that once a solution exists, there will also exist a monotone nondecreasing solution, as discussed above.

The obvious issue with Example 5.10 is that, whereas for \(t > 1\) we get decreasing solutions (thus, nonmonotone, according to our definition of the concept), the solution \(X_t\) is still nondecreasing (and, thus, coincides with \(M_t\)) for all \(t \in [0, 1]\). It is not difficult to modify the example above in such a way that the time threshold \(t = 1\) is replaced by an arbitrarily
small positive time. We mean such that we can replace the time \( t = 1 \) by any smaller time at which a behaviour of maps \( X_t \) changes from increasing to become nonincreasing. The question is, can we construct an example such that a map \( X \), even thought at \( x = 0 \) be the identity (i.e., \( X \) goes to be \( \text{Id} \)), is not nondecreasing as soon as a time \( t \) is positive? However, this natural question can be rewritten as the following

**Question 5.11** Is it true that for any solution \( X \in H^1(0,T;L^2(I)) \) of the flow problem (5.1)–(5.2) with \( X_0 \equiv \text{Id} \), there exists a time horizon \( \bar{t} > 0 \) such that \( X_t \) is nondecreasing for all \( t \in [0,\bar{t}] \)?

As Example 5.12 of the next subsection will show this is false. This is a very nice and highly nontrivial construction, the idea behind it being suggested by A. Bressan. In previous, we lacked an example of a flow that started from the identity map yet became non-non-decreasing instantaneously. Now this is it. In fact, we can construct many examples like Example 5.10 above with a time \( \bar{t} \) for which \( X_t \) is a nondecreasing function of \( x \) for all \( 0 \leq t \leq \bar{t} \). For all those examples we have constructed so far on these times, we have a map \( X \) is nondecreasing in the time interval, but the problem is, can we construct an example such that the map \( X \) become not non-decreasing instantaneously? The answer is we can, but not so easy and this is what we will show you in the next.

### 5.2.3 Examples of a monotonicity and a nonmonotonicity of flows

Let us see now, what can go wrong with Lagrangian flows, which is represented by the problem (5.1)–(5.2), as far as uniqueness of solutions is concerned. In this subsection, we provide some nontrivial and interesting examples which verify uniqueness and nonuniqueness of the flows for several vector fields, including a counterexample showing a vector field for which there exists a nonmonotone flow on any subinterval. Following a suggestion by Alberto Bressan, we have recently constructed this counterexample with a velocity \( v \) and its flow \( X \) such that for any \( \varepsilon > 0 \), there exists a time \( 0 < t_\varepsilon < \varepsilon \) for which the map \( X_{t_\varepsilon} \) is not nondecreasing. Whereas for any positive \( \varepsilon \) is given, \( X_{t_\varepsilon} \) are all nondecreasing from 0 to \( t_\varepsilon \), they can be not nondecreasing (in this case actually they are decreasing after \( t_\varepsilon \)). In this case, we can see in Example 5.10 above that \( \varepsilon = 1 \) at time \( t = 1 \), where the behaviour of the
flow is changed. As we have mentioned above, we can modify this example to make it clear that, instead of the time $t_1 = 1$, we get nonmonotonicity for the flow instantaneously at any small times $t_\varepsilon$.

**Example 5.12 (A counterexample to the monotonicity)** Let $v : [0, 1] \times I \to \mathbb{R}$ be given by $v(t, y) = 2 \text{sgn}(2^{-n/2} - t)\sqrt{1 - 2^{-n}-y}$ for all positive integers $n$, all $1 - 2^{-n+1} < y \leq 1 - 2^{-n}$ and all $t \in [0, 1]$, that is,

$$v(t, y) = \begin{cases} 
2\sqrt{1 - 2^{-n} - y} & \text{if } 1 - 2^{1-n} < y \leq 1 - 2^{-n}, \ 0 \leq t < 2^{-n/2}, \\
-2\sqrt{1 - 2^{-n} - y} & \text{if } 1 - 2^{1-n} < y \leq 1 - 2^{-n}, \ 2^{-n/2} \leq t \leq 1.
\end{cases}$$

If $n \geq 2$ is an integer, define the function $X : [0, 1] \times I \to I$ by

$$X(t, x) = \begin{cases} 
1 - 2^{-n} - (t - \sqrt{1 - 2^{-n} - x})^2 & \text{if } 1 - 2^{1-n} < x < 1 - 2^{-n}, \ 0 \leq t < \sqrt{1 - 2^{-n} - x} \\
1 - 2^{-n} & \text{if } 1 - 2^{1-n} < x < 1 - 2^{-n} - 2^{-n-1}, \ \sqrt{1 - 2^{-n} - x} \leq t \leq 1 \\
1 - 2^{-n} - (t - 2^{-n/2})^2 & \text{if } 1 - 2^{-n} - 2^{-n-1} < x \leq 1 - 2^{-n} - 2^{-n/2} \leq t < 2^{-n/2} \\
1 - 2^{-n} - (t - 2^{-n})^2 & \text{if } 1 - 2^{-n} - 2^{-n-1} < x \leq 1 - 2^{-n}, \ 2^{-n/2} \leq t < 2^{1-n/2} \\
1 - 2^{1-n} & \text{if } 1 - 2^{-n} - 2^{-n-1} < x \leq 1 - 2^{-n}, \ 2^{1-n/2} \leq t \leq 1.
\end{cases}$$

If $n = 1$, we remove the last branch from the above definition and cut $t$ off at 1 in the penultimate branch.

Note that, in this example it is enough to define the velocity $v$ in the space only in the interval $I := (0, 1)$ not on the entire real line $\mathbb{R}$. This basically means it never goes below 0 and never goes above 1. Indeed, the map $v$ is defined for all time $t$ in the time interval $[0, 1]$ and just for all $y$ in the space interval $(0, 1)$, because the map $X$, we constructed above, it is defined here on a bounded domain and takes values only in the space interval $(0, 1)$ as well. In other examples below, $X$ is defined on $(0, 1)$ but does not have to take values in $(0, 1)$, just be any real value. Therefore, when we looking at the flow equation (5.1), the map $v$ does not need to be defined outside $(0, 1)$; that is, it never goes below 0 and never goes above 1. In other words, we can define $v(t, y) \equiv 0$ for $y \in (-\infty, 0) \cup (1, +\infty)$.

Note also that, the initial data $X_0(x) = x$ for all $x \in I$ and the flow problem (5.1)–(5.2) holds in the classical sense for all $x \in I \setminus \{1 - 2^{-n} : n \text{ positive integer} \}$. If $x = 1$, then the differential equation (5.1) is satisfied in the classical sense for $t \in [0, 1]$. If $x = 1 - 2^{-n}$ for
some integer \( n \geq 2 \), then \( X(\cdot, x) \) is continuous in \( t \) on \([0, 1]\), and so is \( \partial_t X(\cdot, x) \) except at 
\( t_n = 2^{1-n/2} \), where it has a jump discontinuity. The differential equation (5.1) is satisfied in 
the classical sense on both sides of \( t_n \), so it is satisfied in the integral sense over the whole 
time interval \([0, 1]\).

Thus, the flow problem (5.1)–(5.2) is satisfied in the integral sense for all \( x \in I \) and 
all \( t \in [0, 1] \). Also, note that on each interval \( I_n := (1 - 2^{-n+1}, 1 - 2^{-n}) \), we have the 
following property: for all \( t \in [2^{-n/2}, 1] \), the map \( X(t, \cdot) \) maps the left half of the interval to 
a single value, namely \( 1 - 2^{-n} \). Then, it maps the right half to a single value as well, namely 
\( 1 - 2^{-n} - (t - 2^{-n/2})^2 \), which lies strictly below \( 1 - 2^{-n} \) for \( t \in (2^{-n/2}, 1] \) (see the interrupted 
line vs the solid line in Figures 5.1, 5.2 below). Thus, \( X(t, \cdot) \) is not Lebesgue a.e. equal to 
a nondecreasing function over \( I_n \) for any integer \( n \geq 2 \) and any \( t \in (2^{-n/2}, 1] \). Since \( 2^{-n/2} \) 
approaches zero as \( n \to +\infty \), we deduce that, for arbitrarily small \( t > 0 \), the function \( X(t, \cdot) \) 
is not Lebesgue a.e. equal to a nondecreasing function over \((0, 1)\).

Since both \( X \) and \( v \) are bounded, we infer \( X \in W^{1,\infty}(0, 1; L^2(I)) \subset H^1(0, T; L^2(I)) \), so 
all the requirements on \( X \) are satisfied. Roughly speaking, the function \( X \) satisfies the flow 
problem (5.1)–(5.2) and all the other assumptions.

Note that, the pictures in Figures 5.1 and 5.2 illustrate how this function of Example 
(5.12) looks like with the initial (particles) data on the interval \((0, 1)\) represented by a vertical 
line and a horizontal line represents time values \( 0 \leq t \leq 1 \). If we imagine that there is a line 
crossing vertically the graph in those figures, that is actually an intercept vertical line starting 
below and intersecting the graph through the interrupted–solid lines. The interrupted ones, 
those the initial particles on \((0, 1)\) start moving in a nice order above the solid ones for 
some time, and then they all accumulate during the lines \( 1 - 2^{-n} \) and go gathering together 
until the specified times \( t = 2^{-n/2} \), where the flow (trajectories) changed its monotonicity 
behaviour. Indeed, at these sequence of times, the trajectories switch their orders such that 
the interrupted ones go below the solid ones (the accumulated lines that keep going together 
straightforwardly). Clearly, the order of particles is not preserved through the time, since 
it is switched after those fixed points in time. We can see the same orders in the beginning 
to those trajectories initially, and then once they reach that times they change order, and 
those times, in which the behaviour of the trajectories is changed, are moving horizontally
along the $t$–axis and ultimately tend to zero whenever $n$ approaches to $+\infty$. This means that, the monotonicity of the flow will become changed obviously from nondecreasing into not nondecreasing. In other words, for each $\varepsilon > 0$, we find a time $t_\varepsilon$ for which the map $X_{t_\varepsilon}$ does not stay nondecreasing.

It is important to make what we pointed out in a previous paragraph clear to readers because this example is not easy to figure out exactly why we choose that velocity map $v$ and why we have such those solutions $X$. In fact, it took a long time until A. Bressan helped us with his suggestion that led to construct this example. So, just to make sure giving an intuition to the readers to understand the nonmonotonicity in this example directly from the picture above, the reader needs to look only at a vertical line crossing the graph.

Finally, note that Example 5.12 can be easily modified to make the flow problem (5.1)–
Figure 5.2: Flow is satisfied in the classical sense for Example 5.12.

(5.2) satisfied in the classical sense for all $t \in (0, 1)$ (see Figure 5.2 above). Indeed, one may replace the fourth branch in the definition of $X$ above by the quadratic equation $u_n(t) = a_n t^2 + b_n t + c_n$ such that $u_n(2^{-n/2}) = 1 - 2^{-n}$, $\partial_t u_n(2^{-n/2}) = 0$ and $u_n(1) = 1 - 2^{1-n}$. For $a_n = -1/(2^{n/2} - 1)^2$, $b_n = 2^{1-n/2}/(2^{n/2} - 1)^2$, $c_n = 1 - 2^{1-n} + [1 - 2^{1-n/2}/(2^{n/2} - 1)^2]$, and $t \in [2^{-n/2}, 1]$, this function will satisfy the differential equation $\partial_t u_n = -\sqrt{4a_n(u_n - c_n) + b_n^2}$. Thus, it suffices to replace the formula in the second branch of the definition of $v$ above by $f(y) = -[2/(2^{n/2} - 1)]\sqrt{1 - 2^{-n} - y}$ if $1 - 2^{1-n} < y \leq 1 - 2^{-n}$ and $2^{-n/2} \leq t \leq 1$.

Not only does Example 5.12 provide an instance where the nonuniqueness of the flow (albeit classical) manifests instantaneously for $t > 0$ (as we have seen there are solutions other than the monotone ones for arbitrarily small positive time), but it also gives instantaneous mass concentrations (Dirac deltas) in the measures $X_t \# \chi$ (as a result of $X_t$ developing flat
portions instantaneously). By “flat portion” we mean the following: for each fixed time \( t \in [0, T] \), the map \( x \rightarrow X_t(x) \), as a function of \( x \), is a constant on some subinterval of \((0, 1)\). Observe that, in each time when we have flat spots on the map \( X_t \), the measure \( \mu_t = X_t#\chi \) has Dirac delta like the one in (5.12). This equivalent to saying that the measure \( \mu_t \) has atoms. The following example shows that this can occur even if the flow is unique (i.e., consisting of monotone nondecreasing maps). In other words, \( X_t \) may develop flat portions as soon as \( t \) becomes positive, i.e. \( X_t \) does not have to remain strictly increasing for small times.

**Example 5.13** Let \( 0 < T \leq 1/2 \) and \( v : [0, T] \times \mathbb{R} \rightarrow \mathbb{R} \) given by

\[
v(t, y) = 0 \quad \text{if} \quad y \leq 0 \quad \text{and} \quad v(t, y) = \frac{y - 1}{1 - t} \quad \text{if} \quad y > 0.
\]

One can prove that

\[
X(t, x) = 0 \quad \text{if} \quad 0 \leq x \leq t \quad \text{and} \quad X(t, x) = \frac{x - t}{1 - t} \quad \text{if} \quad t < x \leq 1
\]

is the only solution to the flow problem (5.1)–(5.2). Note that, the map \( X_t \) develops a flat portion as soon as \( t > 0 \).

Furthermore, as the following example shows, it is also possible that there are two different monotone solutions \( X \) and \( Y \) for the flow problem (5.1)–(5.2). This example is constructed along the lines of the classical \( v(t, y) = 2\sqrt{|y|} \): similarly to the case described in Example 1.3, which is often used to illustrate nonuniqueness for the initial value problem \( \dot{y} = 2\sqrt{|y|}, \; y(0) = 0 \) (in fact, there are infinitely many solutions). This example helps justify why the constraint “\( X_t#\chi \) is prescribed” is used here as a criterion to sort out between solutions. In [6], the author uses this example towards the same goal. The criterion chosen there is different from ours: it is the regularity of the flow, as described earlier in Subsection 2.3.1.

**Example 5.14** Let us consider the autonomous velocity \( v(t, y) = 2\sqrt{|y - 1|}, \; T = 1 \) and
note that the nonunique solutions

\[
X^\varepsilon(t, x) = \begin{cases} 
1 + (t - \sqrt{1 - x})|t - \sqrt{1 - x}| & \text{if } t \in [0, \sqrt{1 - x}], \\
1 & \text{if } t \in [\sqrt{1 - x}, \varepsilon + \sqrt{1 - x}], \\
1 + (t - \varepsilon - \sqrt{1 - x})|t - \varepsilon - \sqrt{1 - x}| & \text{if } t \in [\sqrt{1 - x} + \varepsilon, 1]
\end{cases}
\]

satisfies the flow problem (5.1)–(5.2) for any \(0 \leq \varepsilon < 1\) (see Figure 5.3 below) (the various ranges of \(t\) that do not make sense in the above definition are not to be used: for example, as in Example 1.3, when \(x = 0\), we only use the first branch to get \(X^\varepsilon(t, 0) = 2t - t^2\) for all \(t \in [0, 1]\), or, when \(\varepsilon + \sqrt{1 - x} \geq 1\) the function is defined only on two pieces, namely, \([0, \sqrt{1 - x}]\) and \([\sqrt{1 - x}, 1]\)).

We notice that, the velocity \(v\) given above select the solutions that move immediately away from the singularity \(y = 1\). In fact, it shifted all the “approximating” solutions \(X^\varepsilon\) to the new singularity at \(y = 1\) instead of zero (the case of Example 1.3), because at the value 1 the uniqueness of all these solutions start going wrong: we see the solutions starting increasing from the initial data \(x \in (0, 1)\) to reach the value 1 in time \(t = \sqrt{1 - x}\) and up to that time they are unique, but after that they continue with constant flow 1 until they reach the time \(t = \varepsilon + \sqrt{1 - x}\) for any \(\varepsilon \in [0, 1]\), and then they start changing once again into increasing until the time \(T = 1\).

The maps \(X^\varepsilon_t\) are all nondecreasing (one needs to carefully write \(X^\varepsilon(t, \cdot)\) for fixed \(t\) as a function of \(x\) to check that), yet different for different \(\varepsilon\)—see Figure 5.3 (c). One can easily check that the solution \(X^\varepsilon\) is classical (\(C^1\) in time) and, since it is bounded on \((0, T) \times I\), the flow problem (5.1)–(5.2) provides a bound on \(\dot{X}^\varepsilon\) as well. Thus, \(X^\varepsilon \in H^1(0, T; L^2(I))\) trivially. In terms of measures \(\mu_t := X_t\# \chi\), this means that there are infinitely (in fact, a continuum—one for each \(0 \leq \varepsilon < 1\); see Figure 5.3 (a), (b)) many curves \(\mu \in AC^2(0, 1; \mathcal{P}_2(\mathbb{R}))\), all originating at \(\mu_0 = \chi\) and sharing the same velocity \(v\).

Note that, the pictures (a), (b) in Figure 5.3 below, when we fixing the particle \(x\) and letting the time \(t\) vary this is called the trajectory, are basically showing the trajectory (or the path) \(X_t\) throughout the time for different values of \(\varepsilon\), while the picture (c) shows the union of those different trajectories that is called the flow.
Remark 5.15 The examples we presented above of uniqueness and nonuniqueness all satisfy that the map $X_t$ have flat spots (or portions) as soon as $t > 0$; that is, the measure $\mu_t$ have atoms (atoms mean those Dirac deltas in the measures $X_t \# \chi$).

5.2.4 Connection between $H^1(0, T; L^2(I))$ and $AC^2(0, T; P_2(\mathbb{R}))$ spaces

The concept of monotone rearrangements on the real line, as in Theorem 3.32, makes a connection with the theory of unidimensional optimal transportations that have been discussed in detail in Chapter 4 (see also Chapter 3 for more details of such theory in the high dimensional setting). It has been known by Definition 3.26 that, any Borel map $S$
defined on (in our case) $I$ can be monotonically rearranged over $I$; that is, there exists a nondecreasing map $M : I \to \mathbb{R}$ such that $\chi(S^{-1}(B)) = \chi(M^{-1}(B))$ for all Borel sets $B \subset \mathbb{R}$ (see also Remark 3.27 for an equivalent definition). By $\chi := L^1_{|I}$, we denote the one-dimensional Lebesgue measure restricted to $I$. In other words, there exists a nondecreasing function $M$ such that the Lebesgue measure restricted to $I$ of the preimages of any Borel set through $S$ and $M$ coincide.

**Remark 5.16** By a concept of the one-dimensional Optimal Transport, it is known that there is a bijective correspondence between the closed, convex cone of monotone nondecreasing functions in $L^2(I)$ and the metric space $\mathcal{P}_2(\mathbb{R})$ endowed with the quadratic Wasserstein distance (see, e.g., [70]). This implies [70] (due to the uniqueness of the velocity for a given $AC^2$ curve—see for example Lemma 4.4 of [92] or Theorem 6.5 below) there is a bijective correspondence between paths $M \in H^1(0,T;L^2(I))$ consisting of monotone nondecreasing maps $M_t$ and curves $\mu \in AC^2(0,T;\mathcal{P}_2(\mathbb{R}))$, via $M_t \# \chi = \mu_t$.

Note that, the curve $\mu$ is a time dependent family of measures. We can now combine this fact with the knowledge that each solution to the flow problem (5.1)–(5.2) gives rise to a monotone solution (Theorem 5.9) to state:

**Theorem 5.17** The initial value problem (5.1)–(5.2) admits a solution in $H^1(0,T;L^2(I))$ if and only if $v$ is the velocity associated with some curve $\mu \in AC^2(0,T;\mathcal{P}_2(\mathbb{R}))$.

This may look like a very strong result, yet its practical importance from the point of view of existence for the flow problem (5.1)–(5.2) relies on the fact that one know a priori that a Borel function $v : (0,T) \times I \to \mathbb{R}$ satisfies (2) for $\mu \in AC^2(0,T;\mathcal{P}_2(\mathbb{R}))$ and $v \in L^2(\mu_t)$ for a.e. $t \in (0,T)$. As far as uniqueness for the flow problem (5.1)–(5.2) goes, Example 5.14 shows that, in general, there might be more than one curve whose velocity $v$ is. This accounts for a nonuniqueness “mechanism,” which we now dub multiple-curve (MC) condition on the velocity $v$. Thus, Example 5.14 provides a $v$ satisfying (MC) (or, we could say, “$v$ is (MC)”). If $v$ is not (MC), then we say $v$ is (SC) (single-curve). Examples of $v$ with (SC) are abundant; any continuous $v$ that is also Lipschitz in the $x$-variable would do, as this implies pointwise
existence and uniqueness for \( \dot{y}(t) = v(t, y(t)) \) for any initial \( y = y(0) \). But this regularity is not necessary, as we remark next.

**Remark 5.18** We have seen that the velocity from Example 5.13 produces a single flow map \( X \). If we use it to compute the curve \( \mu = X_\# \chi \), we will get \( \mu_t = t\delta_0 + (1-t)\chi \) (i.e., the convex interpolation between the Dirac mass at 0 and \( \chi \)), and one can check \( \mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R})) \).

Thus, \( v \) is (SC) (even though it is discontinuous) and its (unique) curve develops atoms as soon as \( t > 0 \).

If a velocity \( v \) is (MC), there is no hope for a uniqueness result for the flow problem (5.1)–(5.2), as each curve of probabilities associated with \( v \) will produce its own distinct solution (consisting of the optimal maps). The question of whether the family of optimal maps is the only solution becomes pertinent again once we prescribe the curve associated with \( v \). This is yet another motivation for undertaking the present analysis. The next question is: since \( v \) being (MC) leaves no hope for a uniqueness result for the flow problem (5.1)–(5.2), how about uniqueness of solutions \( X \) for (5.1)–(5.2) under the constraint \( X_t \# \chi = \mu_t \) for \( t \in [0, T] \) (where \( \mu \) is a curve transported by \( v \) via (5.5))? This means that when we impose a constraint on the map \( X_t \); that is, the measure \( \mu_t \) induced via pushforward \( \chi \) is given (i.e. we fix the Eulerian flow), we are not looking on all solutions for the flow problem (5.1)–(5.2) but we are looking only on those pushing forward \( \chi \) to some given measure \( \mu_t \).

By Proposition 4.2 of [70] and Theorem 5.9 above, we know the optimal maps \( M_t \) such that \( M_t \# \chi = \mu_t \) give a solution. We are naturally led to:

**Question 5.19** Let \( \mu \in AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) for some \( T > 0 \) and such that \( \mu_0 = \chi \). Denote by \( v \) the \( L^p(\mu) \) velocity map along the curve \( \mu \). Under what conditions on \( \mu \) do we have that the path \( M \in W^{1,q}(0, T; L^p(I)) \), consisting of the optimal maps such that \( M_t \# \chi = \mu_t \), is the unique solution to the flow problem (5.1)–(5.2), which pushes \( \chi \) forward to \( \mu \)?

**5.2.5 Uniqueness of the flow associated to \( AC^2(0, T; \mathcal{P}_2(\mathbb{R})) \) curves**

We identify here some sufficient conditions on a path \( \mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R})) \) to ensure that the path \( M \in H^1(0, T; L^2(J)) \), consisting of the optimal maps such that \( M_t \# \chi = \mu_t \),
is the unique solution to (5.1)–(5.2), which pushes \( \chi \) forward to \( \mu \). As Example 5.10 refers to a single curve velocity, the nonuniqueness in that case is solely due to having other solutions (other than the one consisting of optimal maps) corresponding to the unique curve of measures associated to the velocity \( v \). Moreover, as we can easily compute, Example 5.13 yields a curve consists of the measures

\[
\mu_t = \begin{cases} 
\frac{1}{1-t} \chi_{(0,1-t)}, & \text{if } t \in [0,1) \\
\delta_0, & \text{if } t = 1 \\
\frac{1}{t-1} \chi_{(1-t,0]}, & \text{if } t \in (1,2].
\end{cases}
\]  

(5.12)

Note that there is a singular measure (Dirac mass) in the middle of the curve. Likewise, in the case of Example 5.14. In terms of the maps \( X_t \), we see that \( X_t \) becomes flat before becoming nondecreasing (i.e., at time \( t = 1 \)). We have not been able to construct an example where \( X_t \) changes behavior from nondecreasing to not nondecreasing without becoming flat first (which translates into \( \mu_t \) having atoms)—see Example 5.12. It is, therefore, natural to ask whether a nonatomic \( \mu_t \) (for all \( t \in [0,T] \)) guarantees the uniqueness we are looking for. Example 5.13 clearly shows, however, that \( \mu \) being nonatomic is not necessary for uniqueness.

To exclude this possibility of existence the singular measure in arbitrary spatial dimension \( d = 1 \), DiPerna and Lions [56], Ambrosio [6] (also, see [41] for a good survey on such problems) have addressed the questions of existence, uniqueness and stability for regular Lagrangian flows, that is, solutions \( X \) of (5.1)–(5.2) such that \( X_t \# \mathcal{L}^1 \ll CL^1 \) for some positive real constant independent of \( t \in [0,T] \) (see Definition 2.8 for the case \( d = 1 \)). In the Sobolev case (there is a similar version if only certain \( BV \) regularity on \( v \) is assumed), the almost (minor improvements are available [41]) state-of-the-art uniqueness result covers only the case \( v \in L^\infty((0,T) \times \mathbb{R}) \cap L^1([0,T];W^{1,p}(\mathbb{R})) \) for some \( p > 1 \) (see Chapter 2 for more details). It is proved in Theorem 2.10 that, if \( v \) is this regular, then its regular Lagrangian flow, if it exists, is unique.

We had originally planned a complete departure from that setting, as our initial goal was to investigate the uniqueness of the Lagrangian flow associated to a given absolutely continuous path of probability measures (see Definitions 5.1 and 6.2). Thus, no explicit
conditions on the velocity $v$ were to be made, even though any assumption on the curve of measures will implicitly reflect on the (unique) velocity associated to it. It turns out that if $d = 1$ and $\chi$ is $L^1$ restricted to $I$, then we can prove a much stronger result for single curve velocities:

**Theorem 5.20** Assume that the Borel map $v : (0, T) \times \mathbb{R} \to \mathbb{R}$ is the velocity of $\mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R}))$ with $\mu_t \ll \chi$ of density $\rho_t$ for $t \in [0, T]$. If $\rho \in L_{loc}^\infty(0, T; L^\infty(\mathbb{R}))$, then (5.1)–(5.2) has a unique solution pushing $\chi$ forward to $\mu$.

**Proof.** We have seen that $\dot{X}_t = v(t, X_t), X(0, \cdot) = X_0$ (with $X_0$ is nondecreasing and $X \in H^1(0, T; L^2(I))$) for some Borel measurable $v : (0, T) \times \mathbb{R} \to \mathbb{R}$ implies that $v$ is the (unique in 1D!) velocity field associated with the curve gives by $t \to \mu_t := X_{t\#}\chi$. Theorem 5.9 asserts that the monotone rearrangements $M_t$ of $X_t$ are such that $M \in H^1(0, T; L^2(I))$ and solves (5.1)–(5.2) as well. By using Theorem 3.32, we have a $\chi$–measure preserving maps $g_t$ for which $X_t = M_t \circ g_t$, we get $\dot{M}(t, g_t(x)) = v(t, M(t, g_t(x))) = \dot{X}(t, x)$. As we can see in Subsection 5.2.4 (also, in an observation in the beginning of the Ambrosio, Gigli, Savare book [17]), we have that $X \in H^1(0, T; L^2(I))$ is equivalent to the $L^2(I)$–norm of $A := (X_{t+h} - X_t)/h - \dot{X}_t$ going to zero as $h$ goes to zero for almost every $t \in (0, T)$. Likewise for the $L^2(I)$–norm $(M_{t+h} - M_t)/h - \dot{M}_t$. But $g_t$ preserves $\chi$, so the latter is equivalent to the $L^2(I)$–norm of $B := (M_{t+h} \circ g_t - M_t \circ g_t)/h - \dot{M}_t \circ g_t$ going to zero. We have already seen that $\dot{M}_t \circ g_t = \dot{X}_t$, so we can write that twice the sum of the squares of the $L^2(I)$–norms of $A$ and $B$ is greater than or equal to the square of the $L^2(I)$–norm of $C := (M_{t+h} \circ g_{t+h} - M_{t+h} \circ g_t)/h$ (because $C = A - B$). Thus, the limit of the latter is zero as $h$ tends to zero. But, if we assume the slopes of all these nondecreasing maps $M_t$ are (at least locally in time) bounded below by a positive constant, we infer that the $L^2(I)$–norm of $(g_{t+h} - g_t)/h$ tends to zero for almost every $t \in (0, T)$. Since $g \in L^2((0, T) \times I)$, we infer (again, by the references above) $g \in H^1(0, T; L^2(I))$ and $\dot{g} \equiv 0$. Thus, $g_t = g_0 = \text{Id}_I$ for almost every $t \in (0, T)$. This means that if $\mu_t$ are functions in $L^\infty(\mathbb{R})$ with their norms locally bounded in time, then $X_t$ must equal $M_t \circ g_0 = M_t$ for almost every $t \in (0, T)$.

Notice that the proof above relies on the *polar decomposition* [34], which is presented in detail in Subsection 3.3.3, of maps $X$ such that $X_{t\#}\chi = \rho_t$; that is, maps $g_t$ such that
$X_t = M_t \circ g_t$ and $g_{t\#} \chi = \chi$. We have been able to prove that $g_t$ is time–independent (equal to the identity map of $I$). As we will see, this theorem is a particular case of Theorem 7.14 when $p = q = 2$ and the proof is very long because here we assumed the function is a regular enough in time to take a derivative and apply the chain rule, but in the reality we don’t have have that in a proof of Theorem 7.14. In that case, $\rho \in AC^2(0, T; \mathcal{P}_2(\mathbb{R})) \cap L^{3}_{loc}(0, T; L^{3}(\mathbb{R}))$ and (5.1)–(5.2) has a unique solution $X \in H^{1}(0, T; L^{2}(\mathbb{I}))$ pushing $\chi$ forward to $\rho$. The idea behind the proof is whenever we use Brenier’s Theorem (Theorem 3.32) to have $\chi$–measure preserving maps $g_t$ such that

$$X(t, x) = M(t, g(t, x)). \quad (5.13)$$

We take the time derivative of both sides to get

$$\dot{X}_t = \dot{M}_t \circ g_t + \partial_x(M_t) \circ g_t \cdot \dot{g}_t. \quad (5.14)$$

But since $M$ solves (5.1)–(5.2) by Theorem 5.9, we can replace $x$ with $g_t$ in $\dot{M}(t, x) = v(t, M(t, x))$ because it is true for a.e. $x$ and $g_t$ is a measure preserving map so it follows that

$$\dot{M}_t \circ g_t = v_t(M_t \circ g_t) = v_t(X_t) = \dot{X}_t \implies \dot{X}_t = \dot{M}_t \circ g_t. \quad (5.15)$$

Comparing (5.14) with (5.15) we get $\partial_x(M_t) \circ g_t \cdot \dot{g}_t = 0$. If $\partial_x M_t$ stays away from zero (thanks to Remark 4.28), then $\dot{g}_t \equiv 0$ for a.e. $t \in (0, T)$. Therefore, $g$ is independent of $t$.

But, $g_0 = \text{Id}_I$ implies $g_t = g_0 = \text{Id}_I$ for a.e. $t \in (0, T)$. Thus, from (5.13) we have $X_t = M_t$.

Note, however, that there is a trade–off between regularity assumptions on $v$, that have been presented in Chapters 1 and L’E-viewpoints, and specifying the pushforward measure. We have discovered that the requirements on $\mu$ can, in fact, be relaxed, but it is not clear to us yet to what extent. It seems likely that the requirement that $\mu_t$ have no atoms for all $t$ is sufficient. In other words, it may be enough to require that $\mu_t$ have no atoms for the monotone solution to (5.1)–(5.2) to be the only solution.

Our main contribution in the next chapters is identifying sufficient conditions on a path $\mu \in AC^q(0, T; \mathcal{P}_p(\mathbb{R}))$, for some $1 \leq p < +\infty$ and $1 \leq q \leq +\infty$, to ensure that the path $M \in W^{1,q}(0, T; L^p(I))$, consisting of the optimal maps such that $M_t \# \chi = \mu_t$, is the unique solution
to (5.1)–(5.2), which pushes $\chi$ forward to $\mu$. First, we shall see in the next chapter that, one can obtain such a curve from (5.1)–(5.2) by defining $\mu_t := X_t#\chi \in AC^q(0, T; \mathcal{P}_p(\mathbb{R}))$; thus, we shall see that if $X_0$ is nondecreasing, then as soon as the flow problem (5.1)–(5.2) admits a solution $X \in W^{1,q}(0, T; L^p(I))$, it automatically admits the solution $M \in W^{1,q}(0, T; L^p(I))$ consisting of the monotone rearrangements of the maps $X_t$. Then, we can prove that there exists a unique Lagrangian flow $X$ of a Borel velocity map $v$ in the sense of Definition 5.1 that pushing $\chi$ forward to $\mu$. This was the main motivation behind our investigation, in the previous subsections, when the flow problem (5.1)–(5.2) has only spatially nondecreasing solutions.
Chapter 6

Lagrangian flows associated to \( AC^q(0, T; \mathcal{P}^{ac}_p(\mathbb{R})) \) curves

In this chapter, we are going to discuss and describe a different look to the Lagrangian viewpoint: the Lagrangian flow of absolutely continuous curves in the \( p^{th} \)-Wasserstein space on the real line (see Section 4.2). We recall that \( \mathcal{P}_p(\mathbb{R}) \) denotes the space of Borel probability measures with finite moment of order \( p \) on \( \mathbb{R} \). The family of absolutely continuous curves in \( \mathcal{P}_p(\mathbb{R}) \), \( 1 \leq p < +\infty \) is central to our approach. The \( p^{th} \)-Wasserstein space \( (\mathcal{P}_p(\mathbb{R}), W_p) \) endowed with the \( L^p \)-Wasserstein distance defined by (4.17) is a Polish space, due to Theorem 4.52, on which we define absolutely continuous curves by Definition 6.2 below. As already observed in the previous chapter: In the Lagrangian description, the flow is described by a function \( X(t, x) \) giving the position of the parcel labeled \( x \) at time \( t \) and the flow velocity is represented by a function \( v(t, y(t)) \). The function \( v \) of position \( y \) and time \( t \) associated with the absolutely continuous curves with finite \( p \)-energy \( E_p(v) := \int_0^T \int_{\mathbb{R}} |v_t(x)|^p \, d\mu_t(x) \, dt < +\infty \), for some \( 1 < p < +\infty \), represents the Eulerian description of fluid flow. The two descriptions on \( (0, T) \times \mathbb{R} \) are related by the differential equation \( \partial_t X(t, x) = v(t, X(t, x)) \), because both sides describe the velocity of the parcel labeled \( x \) at time \( t \). The motivation for our study of this kind of fluid flows that are associated with such absolutely continuous curves on the real line and a brief description of the results was illustrated in Chapter 5 (see also, [21]).

In Section 6.1, we start our discussion of the theory with the velocity vector fields as-
associated to \( L^q \)-absolutely continuous curves in \( \mathcal{P}_p(\mathbb{R}) \). In Subsection 6.1.1, we recall only the definition of absolutely continuous curves in \( AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) and some related results of [17], [92]. Subsection 6.1.2 is devoted to the existence and uniqueness of the velocity fields associated to \( AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) curve. In Subsection 6.1.3, we introduce the concept of \( L^1 \)-velocity of absolutely continuous curves and illustrate by examples how it may not exist in the case \( p = 1 \). Section 6.2.2 is devoted to the Lagrangian flows in the \( W^{1,q}(0, T; L^p(I)) \) space. In Subsection 6.2.1, we illustrate the link between the spaces \( AC^q(0, T; L^p(I)) \) and \( W^{1,q}(0, T; L^p(I)) \). Subsection 6.2.2 is devoted to the connection between the time–Sobolev spaces \( W^{1,q}(0, T; L^p(I)) \) and the class \( AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) of the absolutely continuous curves.

In Section 6.3, we give precise statements of this chapter about existence of the optimal flow maps associated to the \( L^q \)-absolutely continuous curves in \( \mathcal{P}_p(\mathbb{R}) \). Subsection 6.3.1 is devoted to the relationship between the spaces \( AC^q(0, T; L^p(I)) \) and \( AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) in the sense of optimal transportations. In Subsections 6.3.2 and 6.3.3, we state and prove our main results (Theorems 6.33 and 6.35), which show that, quite generally, \( AC \) paths admit Lagrangian descriptions provided by the family of optimal maps consisting of the a time–independent monotone rearrangements of the Lagrangian flow, together with some important corollaries and detailed proofs. Of course, these optimal maps will be defined as the generalized inverses of the cumulative distribution functions of the probability measures on the curve.

6.1 Velocity fields associated to \( AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) curves

In this section, we begin by briefly recalling the definition of some objects essential to our presentation, such as the \( L^p \)-Wasserstein distance and the \( p \)th–Wasserstein space, \( L^q \)-absolutely continuous curves of probability measures on the real line, generalized inverse etc. Then a necessary and sufficient condition on the existence and uniqueness of a Borel velocity field on the real line is proved in order for the Lagrangian flow associated to path of measures possess the \( AC \)-regularity.
6.1.1 Theory of $L^q$–absolutely continuous curves in $P_p(\mathbb{R})$

If $1 \leq p < +\infty$ and $d \geq 1$ is an integer, we denote by $P_p(\mathbb{R}^d)$ the $p$–Wasserstein space on $\mathbb{R}^d$; that is, the set of Borel probabilities on $\mathbb{R}^d$ with finite $p^{th}$–moment order and endowed with the $p$–Wasserstein metric (it can be defined equivalently by (6.16))

$$W_p(\mu, \nu) := \inf \left\{ \left( \mathbb{E}[|X - Y|^p] \right)^{\frac{1}{p}} : \text{law}(X) = \mu, \text{law}(Y) = \nu \right\},$$

where $\mathbb{E}[Z]$ denotes the expected value of a random variable $Z$ and the infimum is taken over all joint distributions of the random variables $X$ and $Y$ with marginals $\mu$ and $\nu$, respectively.

**Remark 6.1** Observe that, thanks to Corollary 3.8, the $p$–Kantorovitch–Rubinstein–Wasserstein distance between Borel probability measures with finite $p$–moment could be defined as

$$W_p^p(\mu, \nu) := \inf \int_{\Omega} |X(\omega) - Y(\omega)|^p d\mathbb{P},$$

where $(\Omega, \mathbb{P})$ probability space and $X, Y \in L^p((\Omega, \mathbb{P}); \mathbb{R}^d)$ such that $X\#\mathbb{P} = \mu$, $Y\#\mathbb{P} = \nu$. For instance, if we take $\Omega = [0, 1]$, then $\mathcal{A}$ is the $\sigma$–algebra of Borel sets on $\Omega$, and $\mathbb{P}$ is the Lebesgue measure on $[0, 1]$. Here, any number between 0 and 1 can be chosen at random, uniformly. In this case, the open intervals of the form $(a, b)$, where $0 < a < b < 1$, could be taken as the generator sets. Each such set can be ascribed the probability of $\mathbb{P}((a, b)) = (b-a)$, which generates the Lebesgue measure on $[0, 1]$, and the Borel $\sigma$–algebra on $\Omega$.

Also, in this thesis $P_p^{ac}(\mathbb{R}^d)$ stands for the set of all Borel probability measures $\mu \in P_p(\mathbb{R}^d)$ which are absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^d$. Let us begin by recalling the definition of $AC^q(0, T, P_p(\mathbb{R}^d))$ (see Appendix A.5.4):

**Definition 6.2** If $1 \leq p < +\infty$ and $1 \leq q \leq +\infty$, a path $[0, T] \ni t \mapsto \mu_t \in P_p(\mathbb{R}^d)$ is said to lie in $AC^q(0, T, P_p(\mathbb{R}^d))$ provided that there exists $\beta \in L^q(0, T)$ such that

$$W_p(\mu_s, \mu_t) \leq \int_s^t \beta(\tau) d\tau \text{ for all } 0 \leq s \leq t \leq T.$$
Moreover, the metric derivative of such a curve is defined as

$$|\mu'(t)| = \lim_{s \to t} \frac{W_p(\mu_s, \mu_t)}{|s - t|} \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, T).$$

(6.1)

It is proved in the book [17] of Ambrosio, Gigli and Savaré that, if \( \rho \in AC^q(0, T; \mathcal{P}_p(\mathbb{R}^d)) \) for some \( 1 \leq q \leq +\infty \) and \( 1 < p < +\infty \), then any such curve admits a Borel velocity map \( \mathbf{v} : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d \) of minimal norm in the sense that:

$$\mathbf{v}(t, \cdot) \in L^p(\mu(t, \cdot); \mathbb{R}^d) \quad \text{for a.e. } t \in (0, T), \quad (0, T) \ni t \mapsto \|\mathbf{v}(t, \cdot)\|_{L^p(\mu(t, \cdot); \mathbb{R}^d)} \in L^q(0, T).$$

(6.2)

$$\int_0^T \|\mathbf{v}(t, \cdot)\|_{L^p(\mu(t, \cdot); \mathbb{R}^d)}^q dt \text{ is minimal among all velocities satisfying (6.2), (6.3).} \quad (6.4)$$

Remark 6.3 It is known that (see [17] Theorem 8.3.1) among all Borel fields \( \mathbf{w} : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d \) satisfying (1.6) and \( \mathbf{w} \in L^p(\mu_t; \mathbb{R}^d) \) for \( \mathcal{L}^1\)-a.e. \( t \in (0, T) \), \( \mathbf{w} \) is the only one with minimal \( L^p \)-norm for a.e. \( t \in (0, T) \). It also satisfies \( \|\mathbf{w}_t\|_{L^p(\mu_t; \mathbb{R}^d)} = |\mu'_t|(t) \) for a.e. \( t \in (0, T) \). We refer to \( \mathbf{v} \) as the tangent velocity field at \( \mu_t \), or the “velocity of minimal norm” associated to the path \( \mu \), as it minimizes \( \int_0^T \|\mathbf{w}(t, \cdot)\|_{L^p(\mu(t, \cdot); \mathbb{R}^d)}^q dt \) for \( \mathcal{L}^1\)-a.e. \( t \in (0, T) \) among all Borel maps \( \mathbf{w} \) that pair up with \( \mu \) to satisfy (6.2) and (6.3). Moreover, it is showed in [17] that, this velocity of minimal norm is unique, in the sense that if \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) satisfy (6.2), (6.3), and (6.4) above for a.e. \( t \in (0, T) \), then \( \mathbf{v}_1(t, \cdot) \equiv \mathbf{v}_2(t, \cdot) \mu_t\text{-a.e.} \). This means that one can choose a velocity associated to the path \( \mu \in AC^q(0, T; \mathcal{P}_p(\mathbb{R}^d)) \) so that it is uniquely determined by the minimal norm property, i.e. if \( \mathbf{w} \) is another velocity associated to the path, then for a.e. \( t \in (0, T) \), \( \|\mathbf{v}_t\|_{L^p(\mu_t; \mathbb{R}^d)} \leq \|\mathbf{w}_t\|_{L^p(\mu_t; \mathbb{R}^d)} \) and \( \mathbf{v}_t \in T_{\mu_t}.\mathcal{P}_p(\mathbb{R}^d) \).

Let us now comment on a motivation that made us look into the general problem of existence for such velocities associated to a given \( AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) curve. The conditions stated in the thesis are, essentially, not directly on the vector field \( v \) but on the (orbit) curve \( \mu_t \) of measures in \( \mathcal{P}(\mathbb{R}) \). It is stated in several places along the thesis (in particular, Remark 6.10 and the statement of Theorem 6.16 (1)) that such a curve in \( AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) determines \( v \) uniquely \( \mu \) a.e.. Indeed, this is an essential point of this dissertation, and all
the results are based on this. However, we start with citing some references and a proof before using this concept in the manuscript. In Remark 6.10, this uniqueness is referred to the valuable reference [92] by Nguyen and Tudorascu. In that reference there is only such a result for \( p = q = 2 \), and even there uniqueness is referred to a minimal norm velocity as in the book of Ambrosio, Gigli, and Savaré [17] on gradient flow. The authors of [92] showed that the minimal norm assumption is, in fact, redundant in the case of one–dimension and they proved that the velocity \( v \) associated to the path \( \mu \in AC^2(0, T; \mathcal{P}_2(\mathbb{R})) \) is unique in the sense mentioned above. We think this point must be cleared, in particular since it is certainly not true, not only in higher dimension but also in dimension one in the circle \( S^1 \).

**Example 6.4** If \( \mu_t(dx) = \rho(x, t)dx \) on \( S^1 \) and \( v \) is a transporting velocity in \( L^\infty(0, T; L^p(\rho; S^1)) \), then \( v + c/\rho(x, t) \) is also transporting in \( L^\infty(0, T; L^p(\rho; S^1)) \) for any \( c \in \mathbb{R} \) if \( 1/\rho \in L^\infty(0, T; L^p(\rho; S^1)) \), that is \( \rho^{1-p} \in L^\infty(0, T; L^1(S^1)) \) (See also Wolansky [118]). Thus, the velocity field is not unique in that case.

It is conceivable that the situation is different on \( \mathbb{R} \), but even for that uniqueness (and existence) of minimal velocity is not clear (at least for a reader) in case \( p = 1 \) (see Remark 6.10). We think, therefore, that this point must be cleared before going to present any results regarding this thesis.

### 6.1.2 Existence and uniqueness of the velocity fields on \( \mathbb{R} \)

As we have pointed out above, the Ambrosio, Gigli and Savaré theory [17] only ensures existence for such an object when \( 1 < p < +\infty \). These authors point to some lecture notes by Ambrosio [4], where it is argued that \( p = 1 \) only guarantees the existence of a Radon (signed) measure \( E \) in time–space such that \( \partial_t \rho + \partial_y E = 0 \). It was not known if \( E \) should disintegrate as \( v(t, y)\rho(t, dy)dt \) if \( p = 1 \). However, there was still the possibility that the special case \( d = 1 \) and the fact that \( \rho_t \in \mathcal{P}^{ac}_1(\mathbb{R}) \) (i.e. the measures \( \rho_t \) are absolutely continuous with respect to the Lebesgue measure) for all \( t \in [0, T] \) would carry the existence result for \( v \) over to the case \( p = 1 \). The absolute continuity assumption on \( \rho_t \) with respect to the Lebesgue measure is a restriction present through the whole thesis, as the main uniqueness results, namely Theorem 7.5 and Theorem 7.14, both assume it. In any case, we clarify by Theorem 6.5 below the
existence and uniqueness of a velocity field associated with a path \( \mu \in AC^q(0,T;P_p(\mathbb{R})) \), including the case \( q = p = 1 \).

In the following we clarify the existence and uniqueness of a velocity field associated with a path \( \mu \in AC^q(0,T;P_p(\mathbb{R}^d)) \), including the case \( p = q = 1 \). We start now by generalizing the result of [92] from \( AC^2(0,T;P_2(\mathbb{R})) \) to \( AC^1(0,T;P_1(\mathbb{R})) \). Note first that \( AC^q(0,T;P_p(\mathbb{R}^d)) \subset AC^1(0,T;P_1(\mathbb{R}^d)) \) for all \( 1 < p < +\infty \) and \( 1 < q \leq +\infty \). More can actually be said if \( d = 1 \); the proof of the theorem below is borrowed, with minor modifications, from [92]. It shows that, on the real line, there is at most one integrable (in the sense specified below) velocity, the minimality condition on its norm being redundant.

**Theorem 6.5** Consider a path \( \mu \in AC^1(0,T;P_1(\mathbb{R})) \) for some \( 0 < T < +\infty \). Then, there exists at most one Borel velocity \( v \) for \( \mu \) such that \( v \in L^1(\mu) \) (as a function of both time and space) for a.e. \( t \in (0,T) \). More precisely, if \( v_1, v_2 : (0,T) \times \mathbb{R} \to \mathbb{R} \) are Borel maps such that \( v_i \in L^1(\mu) \) for \( i = 1,2 \), and such that both \( (\mu,v_1) \) and \( (\mu,v_2) \) satisfy (5.5) in the sense of distributions, then for Lebesgue a.e. \( t \in (0,T) \) we have \( v_1(t,\cdot) \equiv v_2(t,\cdot) \) in the \( \mu(t,\cdot)-\text{a.e.} \) sense.

**Proof.** By subtracting (5.5) written for both \( (\mu,v_1) \) and \( (\mu,v_2) \), and by taking test functions \( \varphi(t,y) = \xi(t)\zeta(y) \), the equations above readily yield

\[
\int_{\mathbb{R}} u(t,y)\zeta'(y) \mu(t,dy) = 0 \quad \text{for a.e. } t \in (0,T) \text{ and any } \zeta \in C^1_c(\mathbb{R}),
\]

where \( u := v_1 - v_2 \). Fix \( \varepsilon > 0 \) and \( \phi \in C_c(\mathbb{R}) \). If \( \phi = 0 \) on \([R, +\infty)\), consider, for each natural number \( n > R \), the function

\[
\Phi_n(y) := \begin{cases} 
\int_{-\infty}^{y} \phi(z) \, dz & \text{if } y < n, \\
\omega(y - n) & \text{if } n \leq y \leq n + 1, \\
0 & \text{if } y > n + 1
\end{cases} \quad (6.5)
\]

where \( \omega \in C^1[0,1] \) such that \( \omega(0) = \int_{-\infty}^{R} \phi(z) \, dz, \omega(1) = 0 \) and \( \omega'(0) = 0 = \omega'(1) \). Clearly,
\( \Phi_n \in C^1_c(\mathbb{R}) \). Thus,

\[
\int_{\mathbb{R}} u(t, y) \phi(y) \mu(t, dy) + \int_n^{n+1} u(t, y) \omega'(y - n) \mu(t, dy) = 0 \quad \text{for a.e. } t \in (0, T).
\]

We have \( |\omega'(y - n)| \leq \|\omega'\|_{L^\infty(0,1)} =: C \) for all \( n > R \) and all \( y \in (n, n+1) \). Since \( u(t, \cdot) \in L^1(\mu(t, \cdot)) \) and \( \mu(t, \cdot) \) is a Borel probability for Lebesgue a.e. \( t \in (0, T) \), we conclude that for such \( t \) and any \( \varepsilon > 0 \) we have

\[
\left| \int_{\mathbb{R}} u(t, y) \phi(y) \mu(t, dy) \right| \leq \varepsilon
\]

if \( n \) is sufficiently large. Due to the arbitrariness of \( \varepsilon \) and \( \phi \), the proof is concluded. \( \square \)

The following is an obvious consequence of Theorem 6.5 and the results of [17] mentioned above (see Theorem 8.3.1 of [17] for more details).

**Corollary 6.6** Let \( 1 \leq p < +\infty \) and \( 1 \leq q \leq +\infty \) and \( \mu \in AC^q(0, T, \mathcal{P}_p(\mathbb{R})) \) be given. Then, there exists at most one Borel map \( v : (0, T) \times \mathbb{R} \rightarrow \mathbb{R} \) such that \((\mu, v)\) satisfies (6.2) (with \( d = 1 \)) and

\[
v \in L^1(\mu), \quad \text{i.e.} \quad \int_0^T \int_{\mathbb{R}} |v(t, y)| \mu(t, dy) dt < +\infty. \quad (6.6)
\]

### 6.1.3 The \( L^1 \)-velocity of absolutely continuous curves in \( \mathcal{P}_p(\mathbb{R}) \)

The uniqueness result above of Theorem 6.5 for the velocity vector field on \( \mathbb{R} \) enables us to make the following important definition:

**Definition 6.7 (The \( L^1 \)-velocity field)** Let \( 1 \leq p < +\infty \) and \( 1 \leq q \leq +\infty \) and \( \mu \in AC^q(0, T, \mathcal{P}_p(\mathbb{R})) \) be given. If it exists, the Borel map \( v : (0, T) \times \mathbb{R} \rightarrow \mathbb{R} \) such that (5.5) and (6.6) are satisfied is called the \( L^1 \)-velocity associated to \( \mu \).

Notice that, the (mostly) uniqueness of the \( L^1 \)-velocity (when it exists!) associated to a given \( AC^q(0, T, \mathcal{P}_p(\mathbb{R})) \) curve was carefully argued in Theorem 6.5 above, which simply extends the earlier result by Nguyen and Tudorascu [92] from \( p = q = 2 \) to the general case \( 1 \leq p < +\infty, 1 \leq q \leq +\infty \).
We now provide some nontrivial and interesting examples which verify the $L^1$–velocity assumption is, in fact, neither redundant nor vacuous in the case $p = 1$. We have an example when $p = 1$ and $v$ exists, and another one when $p = 1$ and $v$ does not exist. We first recall that the measure $\mu$ is absolutely continuous with respect to the density $\rho$, i.e. $\mu \ll \rho$, so by Radon–Nikodym theorem we have $d\mu = \rho dL^1$. If $p = 1$, then even when $d = 1$ and $\rho(t, \cdot) \in P_{1}^{ac}(\mathbb{R})$ for all $t \in (0, T)$, a velocity $v$ satisfying (5.5) and (6.6) may not exist.

**Example 6.8** Let $M : (0, 1) \times (0, 1) \to \mathbb{R}$ be the family of optimal maps given by:

$$M(t, x) = \begin{cases} x & \text{if } x \in [0, 1-t), \\ 1+x & \text{if } x \in [1-t, 1] \end{cases}$$

for all $t \in (0, 1)$. Also, $M(0, x) = x$ and $M(1, x) = 1 + x$ for all $x \in [0, 1]$. Then we can easily compute the curve $\rho(t, \cdot) = M_t \# \chi$ to obtain

$$\rho(t, \cdot) = \chi_{[0, 1-t]} + \chi_{[2-t, 2]},$$

which shows that $\rho(t, \cdot) \in P_{p}^{ac}(\mathbb{R})$ for all $p \geq 1$ and for all $t \in [0, 1]$. However, for all $0 \leq s \leq t \leq 1$

$$W_p(\rho_s, \rho_t) = \left( \int_0^1 |M(s, x) - M(t, x)|^p \, dx \right)^{\frac{1}{p}} = \left( \int_{1-t}^{1-s} 1 \, dx \right)^{\frac{1}{p}} = |t-s|^{\frac{1}{p}},$$

and this is bounded by $\int_s^t \beta(\tau) \, d\tau$ for some $\beta \in L^1(0, 1)$ if and only if $p = 1$ (in which case, we may take $\beta \equiv 1 \in L^\infty(0, 1)$). Thus, $\rho \in AC^\infty(0, 1; P_{1}^{ac}(\mathbb{R}))$, but $\rho \notin AC^q(0, 1; P_{p}(\mathbb{R}))$ for any $1 < p < +\infty$ and any $1 \leq q \leq +\infty$.

Next, assume that the $L^1$–velocity $v$ associated to $\rho$ exists. Then, for all $\varphi \in C^1_c(\mathbb{R})$, the function $t \mapsto \int_{\mathbb{R}} \varphi(y) \rho(t, dy)$ is absolutely continuous on $[0, 1]$ and

$$\frac{d}{dt} \int_{\mathbb{R}} \varphi(y) \rho(t, dy) = \int_{\mathbb{R}} v(t, y) \varphi'(y) \rho(t, dy) \quad \text{at a.e. } t \in (0, 1),$$
Chapter 6. Flows associated to $AC^q(0,T;\mathcal{P}_p(\mathbb{R}))$ curves

that is,

$$-\varphi(1-t) + \varphi(2-t) = \int_0^{1-t} v(t,y)\varphi'(y) \, dy + \int_{2-t}^2 v(t,y)\varphi'(y) \, dy$$

for a.e. $t \in [0,1]$.

Take $\varphi \in C^1_c(\mathbb{R})$ such that $\varphi \equiv 1$ on $[0,5/8]$ and $\varphi \equiv 2$ on $[11/8,2]$. Then, the above equality must be satisfied, in particular, at a.e. $t \in (3/8,5/8)$; this yields $1 = 0$, a contradiction. In conclusion, we have produced an example of a curve lying in the “most regular” subset of the $AC^q(0,1;\mathcal{P}_1(\mathbb{R}))$ families of curves (namely, $AC^\infty(0,1;\mathcal{P}_1^{ac}(\mathbb{R}))$), and yet whose $L^1$–velocity does not exist.

Notice that, Example 6.8 shows that, even curves with the best (in our context) time regularity ($q = +\infty$; Lipschitz curves) will not, in general, admit $L^1$–velocities; indeed, we constructed a curve in $AC^\infty(0,T;\mathcal{P}_1^{ac}(\mathbb{R}))$ for which the $L^1$–velocity does not exist (therefore, this curve will not lie in $AC^1(0,T;\mathcal{P}_p(\mathbb{R}))$ for any $1 < p < +\infty$). On the other hand, the next example provides a curve which lies in $AC^1(0,T;\mathcal{P}_1(\mathbb{R}))$ but not in $AC^q(0,T;\mathcal{P}_p(\mathbb{R}))$ for any $1 < p < +\infty$ and any $1 \leq q < +\infty$; yet for this curve the associated $L^1$–velocity exists.

**Example 6.9** Let $f(z) = (1 - \ln z)^{-1}$, so that $f'(z) = z^{-1}(1 - \ln z)^{-2}$ for $z \in (0,1)$. Then, $f \in L^\infty(0,1)$ and $f' \in L^p(0,1)$ if and only if $p = 1$, so that $f \in W^{1,1}(0,1)$ but $f \notin W^{1,p}(0,1)$ for any $p > 1$. Set $g(x) := \min\{-f'(x),-e/4\}$ and note that $g$ is continuous on $(0,1]$, increasing on $(0,e^{-1})$, and constant on $[e^{-1},1]$. Just like $f'$, we have $g \in L^p(0,1)$ if and only if $p = 1$. Finally, let $M(t,x) := f(t)g(x)$ for all $(t,x) \in [0,1] \times (0,1)$ ($f(0) = 0$ in the right–limit sense) to see that the curve

$$[0,T] \ni t \mapsto \rho(t,\cdot) =: M(t,\cdot) \# \chi$$

lies in $AC^1(0,1;\mathcal{P}_1(\mathbb{R}))$ but not in $AC^q(0,1;\mathcal{P}_p(\mathbb{R}))$ for any $1 \leq p < +\infty$ and any $1 < q \leq +\infty$. Furthermore, the flat portions in the graphs of $M(t,\cdot)$ yield Dirac masses in the measures $\rho(t,\cdot)$ for all $t \in [0,T]$ (while the increasing portions show that these measures are not purely discrete). So, $\rho \notin AC^1(0,1;\mathcal{P}_1^{ac}(\mathbb{R}))$. However, in spite of its very basic $AC^1(0,1;\mathcal{P}_1(\mathbb{R}))$ regularity, one can easily see that $v(t,y) = f'(t)y/f(t)$ if $t \in (0,1]$ is the
L\(^1\)-velocity of \(\rho\). To recapitulate, here we have a curve with no better than \(AC^1(0,1;P_1(\mathbb{R}))\) regularity, but for which the \(L^1\)-velocity exists.

**Remark 6.10** We have included the condition “the \(L^1\)-velocity \(v\) is assumed to exist if \(p = 1\)” in the upcoming statements of Theorem 6.33, Corollary 6.34 and Theorem 7.14, whenever the results apply to \(AC^1(0,1;P_{1ac}(\mathbb{R}))\) curves with their velocities. Examples 6.8 and 6.9 above show that this condition is neither redundant nor vacuous in the case \(p = 1\). Thus, the same will not be necessary for Theorem 7.5, as its statement relies on the a priori existence (and regularity) of the velocity \(v\).

### 6.2 Lagrangian flows in the \(W^{1,q}(0,T;L^p(I))\) space

We now generalize and develop the results for the flow problem (5.1)–(5.2) of Section 5.2 to assume only Borel regularity on \(v\), and study solutions \(X\) belonging to some time–Sobolev spaces \(W^{1,q}(0,T;L^p(I))\). The reason for this extra–requirement on the object introduced in Definition 5.1 will become clear in this section, where we prove that any map \(X\) as in Definition 5.1 that also satisfies \(X_t\#\chi = \rho_t\) for all \(t \in [0,T]\) for some \(\rho \in AC^q(0,T;P_p(\mathbb{R}))\) must, in fact, lie in \(W^{1,q}(0,T;L^p(I))\). This also means that the time–derivative along paths \(X(\cdot, x) \in W^{1,1}(0,T)\), which is denoted by \(\partial_t X(\cdot, x) \in L^1(0,T)\), coincides \(L^2\)-a.e. with the “functional derivative” \(\dot{X} \in L^q(0,T;L^p(I))\) of \(X\). We mean by that \(\dot{X}\) is the weak derivative of the function \(X\) in the time–Sobolev space. In fact, define the time–Sobolev space \(W^{1,q}(0,T;L^p(I))\) is equivalent to known how to define that \(\dot{X}\), so that is basically the weak derivative (like derivative acting by integration parts with respect to smooth test function). In this case, we say that \(X\) in that Sobolev space with respect to time only has a Sobolev derivative that what we denoted by \(\dot{X}\). This definition turns out to be equivalent to requiring that \(X \in W^{1,q}(0,T;L^p(I))\) satisfy

\[
X(t,x) = X(s,x) + \int_s^t v(\tau, X(\tau,x)) \, d\tau \quad \text{for } L^1\text{-a.e. } x \in I \text{ and every } 0 \leq s \leq t \leq T.
\]

On the other hand, we shall prove that a solution \(X\) of the flow problem (5.1)–(5.2) in the sense of Definition 5.1 belongs to \(W^{1,q}(0,T;L^p(I))\) if and only if \([0,T] \ni t \mapsto X_t\#\chi =: \rho_t\)
belongs to $AC^q(0, T, \mathcal{P}_p(\mathbb{R}))$.

### 6.2.1 The spaces $AC^q(0, T; L^p(I))$ and $W^{1,q}(0, T; L^p(I))$

We recall that a Banach space for which every absolutely continuous function has an integrable weak derivative is said to have the Radon–Nikodým property (see, e.g., [23], [36], [55], [119]). This is one of the few properties of real–valued integrals that does not carry over to Bochner integrals in arbitrary Banach spaces. Any reflexive Banach space has this property but, as an example, the spaces $L^1(0, 1)$ and $L^\infty(0, 1)$ do not. Most of the spaces we use are reflexive, and even if they are not, we do not need an explicit characterization of the weakly differentiable functions. Here we recall only that reflexive Banach space, separable dual space, $L^q(0, T; L^p(I))$ for some $1 < p, q < +\infty$ has the Radon–Nikodým property.

It is well known that a real–valued map $X : (0, T) \to \mathbb{R}$ is absolutely continuous if and only if it is weakly differentiable with integrable derivative; that is, $X$ is $L^1$–a.e. differentiable in $(0, T)$, $\dot{X} \in L^1(0, T)$ and $X(t) - X(s) = \int_s^t \dot{X}(\tau) \, d\tau$ for every $0 \leq s \leq t \leq T$ (see Theorem A.10). This is one of the few properties of real-valued integrals that does not carry over to *Bochner integrals* in arbitrary Banach spaces. For maps with values in a Banach space $L^p(I)$, in general this characterizations is not true (see Example 1.17 below). If $1 < p < +\infty$, the Banach space $L^p(I)$ have the Radon–Nikodým property with respect to a Lebesgue measure on $I := (0, 1)$ then this characterization is still true. A similar result holds in the dual space of a separable Banach space $L^p(I)$, with the difference that the absolutely continuous maps are only weakly*–differentiable and the integrals are weak*–integrals.

**Proposition 6.11** For $1 < p < +\infty$, a Banach space $L^p(I)$ is reflexive (respectively; a dual separable space, say $L^{p'}(I)$ such that $1/p + 1/p' = 1$). Then, a map $X \in AC^q(0, T; L^p(I))$ for $1 \leq q \leq +\infty$ if and only if

1. $X$ is differentiable (resp. weakly*–differentiable) for $L^1$–a.e. $t \in (0, T)$,
2. $\lim_{h \to 0} \frac{X_{t+h} - X_t}{h} =: \dot{X}_t \in L^q(0, T; L^p(I))$, (resp. $\lim_{h \to 0} \frac{X_{t+h} - X_t}{h} =: \dot{X}_t \in L^{q}_{w^*}(0, T; L^p(I))$),
3. $X_t - X_s = \int_s^t \dot{X}_\tau \, d\tau$ for every $0 \leq s \leq t \leq T$. 


Moreover, we have
\[ \|\dot{X}_t\|_{L^p(I)} = |X'(t)| \quad \text{for } L^1\text{-a.e. } t \in (0,T). \tag{6.7} \]

**Proof.** We recall that a Banach space $L^p(I)$, $1 < p < +\infty$ satisfying the Radon–Nikodým property. Assume that $X \in AC^q(0,T;L^p(I))$. Given an interval $[a,b] \subset (0,T)$ we set 
\[ \mu([a,b]) := X_b - X_a \] 
and we extend $\mu$ to a vector measure with bounded total variation defined in $BV(0,T)$ absolutely continuous with respect to the Lebesgue measure. We denote the extension still by $\mu$. By the Radon–Nikodým property there exists a Bochner integrable function $\xi : (0,T) \to L^p(I)$ such that $\mu(B) = \int_B \xi(t) \, dt$, for every $B \in BV(0,T)$. In particular, $X_t - X_s = \int_s^t \xi(\tau) \, d\tau$. Then, recalling Theorem A.9, (i) holds and $\dot{X}_t = \xi(t)$ for $L^1$-a.e. $t \in (0,T)$, and consequently (iii) holds. On the other hand, if we denote by $D_hX_t$ the differential quotient $(X_{t+h} - X_t)/h$, then for every point $t \in (0,T)$ where $\dot{X}_t$ exists, by the continuity of the $L^p$-norm we have $\lim_{h \to 0} \|D_hX_t\|_{L^p(I)} = \lim_{h \to 0} \|D_hX_t\|_{L^p(I)}$ and (6.7) holds.

Now (ii) follows from (6.7), since $|X'(t)| \in L^p(0,T)$. 

Now we consider the case of the dual separable space $L^{p'}(I)$ such that $1/p + 1/p' = 1$. For every $\xi \in L^p(I)$, we define $\psi_\xi : (0,T) \to \mathbb{R}$ by $\psi_\xi(t) := \langle \xi, X_t \rangle$. Clearly, $\psi_\xi \in AC^q(0,T;\mathbb{R})$. Taking a countable dense set $\{\xi_k\}_{k \in \mathbb{N}} \subset L^p(I)$, we can see, without difficult, that for $L^1$–a.e. $t \in (0,T)$ there exists $\psi_{\xi_k}(t)$ and $|X'(t)| \leq \lim_{h \to 0} \|D_hX_t\|_{L^p(I)} < +\infty$. For every $t$ as above we define the continuous linear functional $F_t : L^p(I) \to \mathbb{R}$ by continuous extension of $F_t(\xi_k) := \psi_{\xi_k}(t)$, since $|F_t(\xi_k)| \leq \|\xi_k\|_{L^p(I)} |X'(t)|$. Then, $\|F_t\|_{L^{p'}(I)} \leq |X'(t)|$. We show now that $F_t = \dot{X}_t$. Indeed, for every $\delta > 0$, there exists $\xi_k$ such that $\|\xi_k - \xi\|_{L^p(I)} < \delta$, and
\[
|\langle \xi, D_hX_t - F_t \rangle| \leq |\langle \xi - \xi_k, D_hX_t - F_t \rangle| + |\langle \xi_k, D_hX_t - F_t \rangle| \leq \|\xi_k - \xi\|_{L^p(I)} \left( \|D_hX_t\|_{L^p(I)} + |X'(t)| \right) + |\langle \xi_k, D_hX_t - F_t \rangle|,
\]
from which it follows that $\limsup_{h \to 0} |\langle \xi, D_hX_t - F_t \rangle| \leq 2\delta |X'(t)|$. Then (i) is proved and
\[ \|\dot{X}_t\|_{L^p(I)} \leq |X'(t)| \quad \text{for } L^1\text{-a.e. } t \in (0,T). \tag{6.8} \]

The function $\dot{X}_t : (0,T) \to L^p(I)$ is $L^1$ weakly*–measurable, since for every $\xi \in L^p(I)$, 
\[ \langle \xi, \dot{X} \rangle = \lim_{h \to 0} \langle \xi, D_hX_t \rangle \] 
which is limit of continuous functions. Now (ii) follows from (6.8).
We prove (iii): by the definition of the weak*–integral and the absolute continuity of $\psi \xi$ we have $\langle \xi_k, X_t - X_s \rangle = \psi_{\xi_k}(t) - \psi_{\xi_k}(s) = \int_s^t \psi_{\xi_k}(\tau) \, d\tau = \int_s^t \langle \xi_k, \dot{X}_\tau \rangle \, d\tau = \langle \xi_k, \int_s^t \dot{X}_\tau \, d\tau \rangle$. Finally, observing that $\|D_h X_t\|_{L^p(I)} \leq \left| \int_t^{t+h} \|\dot{X}_\tau\|_{L^p(I)} \, d\tau \right|$, then for every Lebesgue point $t \in (0, T)$ of $\|\dot{X}\|_{L^p(I)}$ and every $t \in (0, T)$ such that $|X'(t)|$ exists, we have $\|\dot{X}_t\|_{L^p(I)} \geq |X'(t)|$ and (6.7) holds by (6.8).

Note that, the integral in (iii) is the Bochner integral (resp. the weak*–integral). According to the above theorem (see also Remark 1.1.3 [17]), the absolutely continuous functions are precisely the ones for which the fundamental theorem of calculus holds. The result of this theorem may be regarded as giving an explicit characterization of absolutely continuous curves in Banach spaces $L^p(I)$ for $1 < p < +\infty$. To prove this result one shows from the definition of absolute continuity that if $X \in AC^q(0, T; L^p(I))$, then $\dot{X}_t := \dot{X}(\cdot, x)$ exists pointwise $L^1$–a.e. $t \in (0, T)$ and is integrable, and if $\dot{X}_t = 0$, then $X_t := X(\cdot, x)$ is constant (see for instance Appendix A.5.3). Then, the function $[0, T] \ni t \mapsto X_t - \int_0^t \dot{X}_s \, ds$ is absolutely continuous with pointwise $L^1$–a.e. $t \in (0, T)$ derivative equal to zero, so the result follows. Thus, it follows from the integral representation in Theorem A.10 that every weakly differentiable function with integrable derivative is absolutely continuous, but it can happen an absolutely continuous time–dependent function is not weakly differentiable.

**Example 6.12** If $T = 1$ and $I := (0, 1)$, then $X : I \to L^1(I)$ is defined by $X_t = \chi(0, t)$. Clearly $X_t$ is Lipschitz continuous since, and therefore absolutely continuous. Nevertheless, the derivative $\dot{X}_t$ does not exist for every $t \in (0, 1)$ since the limit as $h \to 0$ of the difference quotients $D_h X_t := (X_{t+h} - X_t)/h$ does not converge in $L^1(I)$, so by Theorem A.10 $X$ is not weakly differentiable. This is a further proof that $L^1(I)$ does not have the Radon–Nikodym property. On the other hand, we consider the same sequence in $C(\bar{I})^*$, identifying $L^1(I)$ with a subset of $C(\bar{I})^*$. The map $C(\bar{I}) \ni \xi \mapsto \langle \xi, X_t \rangle$ is then defined by $\int_I \chi(0, t)(x) \xi(x) \, dx$. Also in this case, we have that $\|X_t - X_s\|_{C(\bar{I})} \leq |t-s|$ for $s \leq t$ in $I$, and $\langle \xi, D_h X_t \rangle \to \xi(t) = \langle \xi, \delta_t \rangle$ as $h \to 0$, but $D_h X_t$ does not strongly converge.

The following Lemma shows that the spaces $AC^q(0, T; L^p(I))$ are strictly linked to the Sobolev spaces $W^{1,q}(0, T; L^p(I))$. 

Lemma 6.13 Let \(1 < p < +\infty\) and \(1 < q \leq +\infty\). If \(X \in AC^q(0, T; L^p(I))\) then (the equivalence class of) \(X \in W^{1,q}(0, T; L^p(I))\). If \(X \in W^{1,q}(0, T; L^p(I))\) then there exists a unique continuous representative \(\tilde{X} \in C(0, T; L^p(I))\) (in particular \(\tilde{X}_t = X_t\) for \(L^1\)-a.e. \(t \in (0, T)\)). Moreover \(\tilde{X} \in AC^q(0, T; L^p(I))\) and the application \(F : W^{1,q}(0, T; L^p(I)) \to C(0, T; L^p(I))\) defined by \(F|_X = \tilde{X}\) is a Borel map.

Proof. The proof of first assertion can be carried out exactly as in [36] for the well known case of the real–valued maps by using \(|D_h X_t|^q_{L^p(I)} \leq \frac{1}{h} \int_t^{t+h} |X'|^q(s) \, ds\), where \(D_h X_t\) denotes the difference quotients \((X_{t+h} - X_t)/h\).

Now we assume that \(X \in W^{1,q}(0, T; L^p(I))\) and we consider a sequence \(\{x_k\}_{k \in \mathbb{N}}\) dense in \(L^p(I)\). Defining \(X^k_t := \|X_t - x_k\|_{L^p(I)}\), the triangular inequality implies

\[
|X^k_{t+h} - X^k_t| \leq \|X_{t+h} - X_t\|_{L^p(I)}, \tag{6.9}
\]

where \(h\) is sufficiently small such that \(t+h \in (0, T)\). The fact that \(X \in W^{1,q}(0, T; L^p(I))\) and \(1 \leq p < +\infty, 1 < q \leq +\infty\) implies that \(\|X^k_t\|_{L^p(I)} \in W^{1,q}(0, T)\) again for [36] Proposition 8.3. Hence there exist \(\tilde{X}^k_t\) absolutely continuous such that \(\tilde{X}^k_t = X^k_t\) a.e. and \(\dot{\tilde{X}}^k_t\) is a.e. differentiable. We introduce the negligible set

\[
N = \bigcup_{k \in \mathbb{N}} \left\{t \in (0, T) : \tilde{X}^k_t \neq X^k_t \text{ and } \dot{\tilde{X}}^k_t \text{ does not exist}\right\}
\]

and we define \(\beta(t) := \sup_{k \in \mathbb{N}} |\dot{\tilde{X}}^k_t|\) for every \(t \in (0, T) \setminus N\). Clearly, by the density of \(\{x_k\}\), we have for every \(s \leq t\) in \((0, T)\setminus N\)

\[
\|X_t - X_s\|_{L^p(I)} = \sup_{k \in \mathbb{N}} |\tilde{X}^k_t - \tilde{X}^k_s| \leq \sup_{k \in \mathbb{N}} \int_s^t |\dot{\tilde{X}}^k_{\tau}| \, d\tau \leq \int_s^t \beta(\tau) \, d\tau. \tag{6.10}
\]

We show that \(\beta \in L^p(I)\). Actually, by (6.9), if \(t \in (0, T) \setminus N\) then

\[
|\dot{\tilde{X}}^k_t| = \lim_{h \to 0} \left| \frac{\tilde{X}^k_{t+h} - \tilde{X}^k_t}{h} \right|_{L^p(I)} \leq \liminf_{h \to 0} \|D_h X_t\|_{L^p(I)}, \tag{6.11}
\]

which implies \(\beta(t) \leq \liminf_{h \to 0} \|D_h X_t\|_{L^p(I)}\). By Fatou’s Lemma and \(X \in W^{1,q}(0, T; L^p(I))\)
we obtain
\[ \int_0^T |\beta(t)|^q(t) \, dt \leq C. \]  
(6.12)

Since (6.10) and Hölder’s inequality show that \( X : (0, T) \setminus N \rightarrow L^p(I) \) is uniformly continuous, thus, by the completeness of \( L^p(I) \), it admits a unique continuous extension \( \tilde{X} : (0, T) \rightarrow L^p(I) \) which also satisfies
\[ \|\tilde{X}_t - \tilde{X}_s\|_{L^p(I)} \leq \int_s^t \beta(\tau) \, d\tau \quad \text{for every } s \leq t \text{ in } (0, T), \]  
(6.13)

and then, for (6.12), \( \tilde{X} \in AC^q(0, T; L^p(I)) \). We have thus proved that \( X \in AC^q(0, T; L^p(I)) \) if and only if \( X \in C(0, T; L^p(I)) \) and \( \sup_{0 < h < T} \int_0^{T-h} \|D_h X_t\|_{L^p(I)}^q \, dt \leq +\infty \).

In order to prove that the extension operator \( F \) is a Borel map, we observe that \( W^{1,q}(0, T; L^p(I)) \) and \( AC^q(0, T; L^p(I)) \) are Borel subsets of \( L^q(0, T; L^p(I)) \) and \( C(0, T; L^p(I)) \) respectively, since the map \( X \mapsto \sup_{0 < h < T} \int_0^{T-h} \|D_h X_t\|_{L^p(I)}^q \, dt \) is lower semi continuous from \( L^q(0, T; L^p(I)) \) to \([0, +\infty]\) and from \( C(0, T; L^p(I)) \) to \([0, +\infty]\). Moreover, \( F \) is an isometry from \( (W^{1,q}(0, T; L^p(I)), \|\cdot\|_{L^q(0,T)}) \) to \( (C(0, T; L^p(I)), \|\cdot\|_{L^q(0,T)}) \) and the thesis follows by observing that the Borel sets of \( (C(0, T; L^p(I)), \|\cdot\|_{L^q(0,T)}) \) coincides with the Borel sets of \( (C(0, T; L^p(I)), \|\cdot\|_{\infty}) \). This last assertion is a general fact: if \( X \) is a separable and complete metric space (Polish space) and \( X_w \) is the same space with an Hausdorff topology weaker than the original, then the Borel sets of \( X \) coincides with the \( X_w \) ones (see for instance; Corollary 2 of [106]).

The following theorem basically states that the \( L^1 \)–velocity field \( v \) that is satisfying the conditions (6.2) and (6.3) is; indeed, the functional time–derivative of the Lagrangian flow \( X \) in \( L^q(0, T; L^p(I)) \), in the sense of Definition 5.1.

**Theorem 6.14** Let \( X \) be a Lagrangian flow of a Borel vector field \( v_t \), in the sense of Definition 5.1, satisfying the \( L^q \)–integrability condition
\[ \int_0^T \|v_t\|_{L^p(\mu_t)}^q \, dt < +\infty, \quad \text{for some } 1 < p < +\infty, 1 < q \leq +\infty. \]  
(6.14)
Then, the velocity field $v_t$ is the time-derivative of $X_t$ in the $L^q$-sense

$$\lim_{h \to 0} \int_0^{T-h} \left( \int_{\mathbb{R}^d} \frac{X_{t+h}(x) - X_t(x)}{h} - v_t(X_t(x)) \right)^q dx \quad dt = 0. \quad (6.15)$$

**Proof.** This is can be easily obtained by observing that the differential quotient $D_h X(\cdot, x) := (X_{t+h}(x) - X_t(x))/h$ is bounded in $L^q(0, T; L^p(I))$ by

$$\int_0^{T-h} \left( \int_I \frac{X_{t+h}(x) - X_t(x)}{h} \right)^q dx \quad dt = \int_0^{T-h} \left( \int_I \frac{v_s(X_s(x)) ds}{h} \right)^q dx \quad dt$$

$$\leq \int_0^{T-h} \left( \int_I \left| v_s(X_s(x)) \right|^p ds \right)^{q/p} dx \quad dt \leq \int_0^{T} \left( \int_I \left| v_t(X_t(x)) \right|^p dx \right)^{q/p} dt$$

$$\leq \int_0^{T} \left( \int_I \left| v_t(x) \right|^p d\mu_t(x) \right)^{q/p} dt < +\infty.$$

Since we already know that $D_h$ is pointwise converging to $v_t \circ X_t$ for $L^1 \times \chi$-a.e. in $(0, T) \times I$, we obtain the strong convergence in $L^q(0, T; L^p(I))$; i.e. (6.15).

Now, we consider $t \mapsto X_t(\cdot)$ and $t \mapsto v_t(X_t(\cdot))$ as maps from $(0, T)$ to $L^p(I)$; (6.15) is then equivalent to

$$\lim_{h \to 0} \int_0^{T-h} \left\| \frac{X_{t+h} - X_t}{h} - v_t(X_t) \right\|^q_{L^p(I)} dt = 0,$$

and it shows that $t \mapsto X_t(\cdot)$ belongs to $AC^q(0, T; L^p(I))$. General results for absolutely continuous maps in reflexive Banach spaces (due to Proposition (6.11)), yield that $X_t$ is differentiable $L^1$-a.e. in $(0, T)$, so that

$$\lim_{h \to 0} \int_I \left| \frac{X_{t+h}(x) - X_t(x)}{h} - v_t(X_t(x)) \right|^p dx = 0 \quad \text{for } L^1 \text{-a.e. } t \in (0, T),$$

hence we have shown the desired thesis. $\square$

As a consequence, since the following *semigroup property* $X_{t+h}(x) = X_h(X_t(x), x)$ holds (due the uniqueness of the Lagrangian flows) and then we obtain

$$\lim_{h \to 0} \frac{X_{t+h}(t, x) - x}{h} = v_t(x) \quad \text{in } L^p(\mu_t) \quad \text{for } L^1 \text{-a.e. } t \in (0, T).$$
6.2.2 The spaces $W^{1,q}(0, T; L^p(I))$ and $AC^q(0, T; P_p(\mathbb{R}))$

It is interesting to consider the relationship between weak differentiability and absolute continuity of the probability measures on the real line. We assume first that the reader is familiar with the concept of the $p$th–Wasserstein distance on the real line (See Section 4.2).

**Lemma 6.15** Let $1 < p < +\infty$ and $1 < q \leq +\infty$. If $X \in W^{1,q}(0, T; L^p(I))$ is a Lagrangian flow associated to $(\rho, v)$ such that a Borel velocity field $v$ satisfying (6.14), then $\rho \in AC^q(0, T; P_p(\mathbb{R}))$. Moreover, the metric derivative $|\rho'|(t)$ exists for $L^1$–a.e. $t \in (0, T)$.

**Proof.** By taking into account (5.1)–(5.2) and (6.14), and using Hölder’s inequality with Fubini’s theorem, we obtain that for every $s, t \in [0, T]$, with $s < t$,

$$
\int_I |X_s(x) - X_t(x)|^p \, dx \leq (t - s)^{p-1} \int_s^t \|v_\tau(x)\|^p \, d\rho_\tau(x) \, d\tau.
$$

Recalling the definition of the $p$th–Wasserstein distance between $\rho_s, \rho_t \in P_p(\mathbb{R})$ that is provided, through the following relation with a minimal transport cost of (KP)

$$
W_p(\rho_s, \rho_t) := \min_{\gamma \in \Gamma(\rho_s, \rho_t)} \left( \int \int_{\mathbb{R}^2} |x - \bar{x}|^p \, d\gamma(x, \bar{x}) \right)^{\frac{1}{p}},
$$

where $\Gamma(\rho_s, \rho_t)$ is the set of Borel probability measures on the product space $\mathbb{R} \times \mathbb{R}$ having first marginal equal to $\rho_s$ and second marginal equal to $\rho_t$ (see Definition 6.16), and using the admissible measure $(X_s(\cdot), X_t(\cdot)) \# \chi =: \gamma \in \Gamma(\rho_s, \rho_t)$ we obtain the inequality

$$
W_p(\rho_s, \rho_t) \leq (t - s)^{p-1} \int_s^t \|v_\tau(x)\|^p \, d\rho_\tau(x) \, d\tau.
$$

(6.17)

This last inequality and (6.14) imply that the curve $t \mapsto \rho_t$ is absolutely continuous in $P_p(\mathbb{R})$. By Lebesgue differentiation theorem (see Theorem A.10), we have the metric derivative of $\rho$

$$
|\rho'|^p(t) := \lim_{s \to t} \frac{W_p^p(\rho_s, \rho_t)}{|t - s|^p} \leq \int_{\mathbb{R}} \|v_t(x)\|^p \, d\rho_t(x) \quad \text{for } L^1\text{–a.e. } t \in (0, T),
$$

(6.18)

that is, $|\rho'|(t) \leq \|v_t\|_{L^p(\rho_t)}$ for $L^1$–a.e. $t \in (0, T)$ and, because of the assumption (6.14), we obtain that the curve $\rho_t$ belongs to the space $AC^q(0, T; P_p(\mathbb{R}))$, i.e. the space of the absolutely continuous curves $\rho_t : [0, T] \to P_p(\mathbb{R})$ such that $|\rho'| \in L^q(0, T)$.

\[ \Box \]
Notice that, the application \( t \mapsto \rho_t \) can be thought as a curve in the metric space \( \mathcal{P}_p(\mathbb{R}) \) of Borel probability measures with finite \( p \)-th moment, endowed with the distance \( W_p \) defined in (6.17). The property (6.18) says that the curve \( t \mapsto \rho_t \) is absolutely continuous, hence the metric derivative \( |\mu'|(t) \) exists \( L^1 \)-a.e. \( t \in (0, T) \), and, for (6.18) and (6.14), the \( q \)-energy of the curve \( \rho \); that is, \( E_q(\rho) := \int_0^T |\rho'|^q(t) \, dt \), is finite.

The theorem we prove next establishes the connection between the regularity of an absolutely continuous curve of probabilities and that of its Lagrangian description. Note that there is no claim at this point that said description be unique; this result applies to any flow map associated to the given curve.

**Theorem 6.16** Let \( 1 \leq p < +\infty \) and \( 1 \leq q \leq +\infty \). If \( X : [0, T] \times I \to \mathbb{R} \) is a Lagrangian flow map associated to the Borel map \( v : [0, T] \times \mathbb{R} \to \mathbb{R} \), then the following are equivalent:

1. \( X \# \chi = \rho \in AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \) and \( v \) is its unique \( L^1 \)-velocity field (or, simply, the velocity field associated to \( \rho \)). In fact,

\[
(0, T) \ni t \mapsto \|v_t\|_{L^p(\rho_t)} \in L^q(0, T).
\] (6.19)

2. \( X \in W^{1,q}(0, T; L^p(I)) \), in which case \( \partial_t X \) coincides a.e. in \((0, T) \times I\) with the functional derivative \( \dot{X} \) of \( X \).

**Proof.** (1) \( \implies \) (2). Note that

\[
\int_0^T \|v_t\|^q_{L^p(\rho_t)} \, dt = \int_0^T \left( \int_I |v(t, X(t, x))|^p \, dx \right)^{\frac{q}{p}} \, dt,
\]

so \( \partial_t X \in L^q(0, T; L^p(I)) \subset L^1((0, T) \times I) \). Now,

\[
|X(t, x)| \leq |X_0(x)| + \int_0^t |v(s, X(s, x))| \, ds
\]

\( \implies \|X(t, \cdot)\|_{L^p(I)} \leq \|X_0(x)\|_{L^p(I)} + \int_0^t \|\partial_t X(s, \cdot)\|_{L^p(I)} \, ds \) for all \( t \in [0, T] \).

Thus, \( X \in L^\infty(0, T; L^p(I)) \subset L^1((0, T) \times I) \). Since \( X(\cdot, x) \in W^{1,1}(0, T) \) for a.e. \( x \in I \), we
have
\[\int_0^T \partial_t \varphi(t,x)X(t,x) \, dt = - \int_0^T \varphi(t,x) \partial_t X(t,x) \, dt \quad \text{for all } \varphi \in C^1_c((0,T) \times I).\]

We use \(X, \partial_t X \in L^1((0,T) \times I)\) to conclude
\[\int_0^T \int_I \partial_t \varphi(t,x)X(t,x) \, dx \, dt = - \int_0^T \int_I \varphi(t,x) \partial_t X(t,x) \, dx \, dt \quad \text{for all } \varphi \in C^1_c((0,T) \times I).\]

Thus, \(X \in W^{1,q}(0,T; L^p(I))\) and \(\partial_t X = \dot{X}\) a.e. in \((0,T) \times I\).

(2) \(\implies\) (1). To prove \(\rho \in AC^q(0,T; \mathcal{P}_p(\mathbb{R}))\), note that the fact that \((X, X)\# \chi\) is a transport plan between \(\rho_s\) and \(\rho_t\) (see the observations of (6.17)) implies
\[W_p(\rho_s, \rho_t) \leq \|X_s - X_t\|_{L^p(I)} \leq \int_s^t \|\dot{X}(\tau, \cdot)\|_{L^p(I)} \, d\tau \quad \text{for all } 0 \leq s \leq t \leq T. \quad (6.20)\]

According to Theorem 8.3.1 [17], the fact that the map \([0,T] \ni t \mapsto \|\dot{X}(t, \cdot)\|_{L^p(I)}\) lies in \(L^q(0,T)\) implies \(\rho \in AC^q(0,T; \mathcal{P}_p(\mathbb{R})) \subset AC^1(0,T; \mathcal{P}_1(\mathbb{R}))\). It is now enough to prove (6.19) (which implies the \(L^1\)-velocity field as in Definition 6.7) and the fact that \((\rho, v)\) satisfies (5.5) as distributions. Property (6.19) follows immediately (via \(X = \rho_t\) and the a.e. identification \(\partial_t X = \dot{X}\)) from
\[\int_0^T \|\dot{X}(t, \cdot)\|^q_{L^p(I)} \, dt = \int_0^T \left(\int_I |v(t,X(t,x))|^p \, dx\right)^{q/p} \, dt = \int_0^T \|v_t\|^q_{L^p(I)} \, dt. \quad (6.21)\]

Since any \(\varphi \in C^1_c(I)\) is Lipschitz, we have that, for a.e. \(x \in I\), the function \([0,T] \ni t \mapsto \varphi(X(t,x))\) lies in \(W^{1,1}(0,T)\) and its a.e. derivative is \(\varphi'(X(t,x))\partial_t X(t,x)\). Or, equivalently, for a.e. \(x \in I\), we have (we use the flow problem (5.1)–(5.2) for the second equality)
\[\int_0^T \xi(t) \varphi(X(t,x)) \, dt = - \int_0^T \xi(t) \varphi'(X(t,x)) \partial_t X(t,x) \, dt = - \int_0^T \xi(t) \varphi'(X(t,x)) v(t,X(t,x)) \, dt \quad \text{for all } \xi \in C^1_c(0,T).\]

But both integrands are in \(L^1((0,T) \times I)\), so by Fubini’s Theorem the above equality can be
integrated in $x$; then we can change the order of integration to get

$$\int_0^T \xi(t) \int_I \varphi(X(t, x)) \, dx \, dt = - \int_0^T \varphi'(X(t, x)) v(t, X(t, x)) \, dx \, dt$$

for all $\xi \in C^1_c(0, T)$.

Since $X_t \# \chi = \rho_t$, the above translates into

$$\int_0^T \xi(t) \int_{\mathbb{R}} \varphi(y) \rho_t(dy) \, dt = - \int_0^T \xi(t) \int_{\mathbb{R}} \varphi'(y) v(t, y) \rho_t(dy) \, dt. \quad (6.22)$$

This proves that $v$ is, indeed, a velocity field for $\rho$, and, according to Theorem 6.5, it is the only one satisfying (5.5) and (6.6).

Note that, the fact that $v$ is the appropriate velocity field, in the sense of Definition 6.7, is proved in Theorem 6.16 above. Indeed, thanks to Corollary 6.6, we have $\rho \in AC^1(0, T; P_1(\mathbb{R}))$, and its metric derivative $|\rho'|(t) = \int_I |\dot{X}(t, x)| \, dx$ for a.e. $t \in (0, T)$.

Furthermore, $v$ is the unique map satisfying (5.5) and the condition (1.7) (with $d = 1$), i.e.

$$v_t \in L^1(\rho_t) \quad \text{for a.e. } t \in (0, T) \quad \text{and} \quad (0, T) \ni t \mapsto \|v_t\|_{L^1(\rho_t)} \in L^1(0, T).$$

**Remark 6.17** Because of the uniqueness of the $L^1$–velocity $v$ (when it exists) for a given $\rho \in AC^1(0, T; P_1(\mathbb{R}))$, every time we refer to a Lagrangian flow map $X$ for $v$ under the constraint $X_t \# \chi = \rho_t$, we may simply call it a Lagrangian flow map associated to $\rho$.

### 6.3 Optimal flow maps in the $W^{1,q}(0, T; L^p(I))$ space

Motivated by work of Gangbo, Nguyen, and Tudorascu [70], proved a surprisingly general flow–map formula which unequivocally links an absolutely continuous curve in the Wasserstein space to the corresponding family of optimal maps pushing forward a given reference measure to each measure on the curve. In this section we prove a similar result for the higher integrability case. Possible applications to a time–independent rearrangement maps for Lagrangian flows are discussed. These time–independent rearrangements are obtained as a generalized inverse of the cumulative distribution function of the Borel probability measure.

As we mentioned earlier in Subsection 5.2.4, there is a connection between the theory
of one–dimensional Optimal Transport and the concept of monotone (nondecreasing) rearrangement of the Lagrangian flows $X_t : I \to \mathbb{R}$. Indeed, there exists a nondecreasing function $M_t : I \to \mathbb{R}$ such that the Lebesgue measure restricted to $I$ of the pre–images of any Borel set through $X_t$ and $M_t$ coincide for every $t \in (0, T)$. This function satisfies

$$M_t(x) = \inf \{ y \in \mathbb{R} : F_t(y) > x \} \text{ for all } x \in (0, 1),$$

where $F_t$ is the cumulative distribution function of the Borel probability measure $\mu := S_\# \chi$; that is, $M_t$ is the generalized inverse of $F_t$. It also turns out that $F_t$ is the generalized inverse of $M_t$.

**Remark 6.18** If $\mu_t$, $t \in (0, T)$ has no atoms, then $F_t \# \mu_t = \chi$ optimally. $M_t$ is strictly increasing and $F_t \circ M_t \equiv \text{Id}$. We also have that if $\mu_t \ll \mathcal{L}^1$, then $M_t \circ F_t \equiv \text{Id}$ on the support of $\mu_t$.

As consequence of Remark 4.45 is the following lemma which shows the equivalency between the class of $L^q$–absolutely continuous (Borel) maps in $L^p(I)$ and the class of $L^q$–absolutely continuous curves $\mathcal{P}_p(\mathbb{R})$. We then exploit the expression of the metric derivatives given in (6.1) to conclude that $|\mu'|(t) \leq |M'|(t)$ for $\mathcal{L}^1$–a.e. $t \in (0, T)$. As a consequence, $\|M'\|_{L^q(0,T)} \geq \|\mu'\|_{L^q(0,T)}$.

**Lemma 6.19** Let $1 < p < +\infty$ and $1 < q \leq +\infty$. Suppose $\mu \in L^q(0, T; \mathcal{P}_p(\mathbb{R}))$, $M \in L^q(0, T; L^p(I))$ are such that $M_t$ is monotone nondecreasing and $M_t \# \chi = \mu_t$. Then, $\mu \in AC^q(0, T; \mathcal{P}_p(\mathbb{R}))$ if and only if $M \in AC^q(0, T; L^p(I))$. In that case $|M'|(t)$ exists if and only if $|\mu'|(t)$ exists. Moreover, both functions coincide where they exist.

**Proof.** The proof is a direct consequence of Remark 4.45 and (6.1). \qed

### 6.3.1 The spaces $AC^q(0, T; L^p(I))$ and $AC^q(0, T; \mathcal{P}_p(\mathbb{R}))$

In the remainder of our study is restricted on the special case $d = 1$, so recall that

$$I = (0, 1), \quad I_T = (0, T) \times I, \quad \chi = \mathcal{L}^1|I, \quad \chi_T = \mathcal{L}^1|_{(0,T)\times I}.$$
One can notice that the next lemmas is still valid if we replace $I$ by an open subset of $\mathbb{R}$. The first lemma of this section recalls (in the particular case when $p = q$) the standard identification of $L^q(0,T; L^p(I))$ and $L^q(I_T)$ (the space of all Lebesgue measurable functions $M$ with respect to measure $\chi_T$). It allows not to distinguish between these two spaces. Given $N$ as in the lemma below, replacing $N$ by $\tilde{N}$ if necessary, we shall always use the convention that $N \equiv \tilde{N}$. Since the proof of the lemma is standard, it will be skipped.

**Lemma 6.20** If $N \in L^q(0,T; L^q(I))$ for some $1 \leq q < +\infty$, then there exists $\tilde{N} \in L^q(I_T)$ such that $\int_0^T dt \int_I N_t(x) \zeta(t,x) dx = \int_{I_T} \tilde{N}(x,t) \zeta(t,x) dx dt$ for all functions $\zeta \in L^q(I_T)$. Furthermore, $t \mapsto \tilde{N}(\cdot,t)$ belongs to $L^q(0,T; L^q(I))$ and for $L^1$–a.e. $t \in (0,T)$, $\tilde{N}(x,t) = N_t(x)$ for $\chi$–a.e. $N \in I$.

The next lemma is elementary and so, its proof will not be given.

**Lemma 6.21** Suppose that $M$ and a sequence $\{M^k\}_{k \in \mathbb{N}}$ belong to $AC^q(0,T; L^p(I))$ for some $1 \leq q \leq +\infty$ and $1 \leq p < +\infty$ such that $\|M^k\|_{AC^q(0,T; L^p(I))}, \|M\|_{AC^q(0,T; L^p(I))} \leq C$ for a constant $C > 0$. Suppose that for each $t \in (0,T)$, $\{M^k_t\}_{k \in \mathbb{N}}$ converges weakly to $M_t$ in $L^p(I)$. Then, $\{M^k\}_{k \in \mathbb{N}}$ converges weakly to $M$ in $L^q(0,T; L^p(I))$ and $\{\dot{M}^k\}_{k \in \mathbb{N}}$ converges weakly to $\dot{M}$ in $L^q(0,T; L^p(I))$.

In the remainder of this work $d = 1$ and $I = (0,1)$. The purpose of the next two lemmas is to show that if $M \in AC^q(0,T; L^p(I))$ and $M_t$ is monotone nondecreasing and right continuous for each $t$, then $(t,x) \mapsto M_t(x)$ is a Borel map. The point is that we do not need to modify $M_t(x)$ on a set of $\mathcal{L}^2$–zero measure to obtain a Borel map.

**Lemma 6.22** Let $a < b$ be two real numbers and let $M \in AC^q(0,T; L^p(a,b))$ for some $1 \leq q \leq +\infty$ and $1 \leq p < +\infty$. Suppose that for each $t$, the function $M_t : (a,b) \rightarrow \mathbb{R}$ is monotone, nondecreasing and continuous. Then, $(t,x) \mapsto M_t(x)$ is continuous on $(0,T) \times I$.

**Proof.** We skip the proof of this lemma since it is an elementary exercise. We give a hint which is based on the following fact on the class of $C^1(a,b)$–convex functions. Suppose $\{\xi_k\}_{k \in \mathbb{N}} \subset C^1(a,b)$ are convex, $\xi \in C^1(I)$ is convex and $\|\xi_k\|_{L^1(a,b)}$ is bounded. Then $\{M^k\}_{k \in \mathbb{N}}$ converges weakly in $L^1_{\text{loc}}(a,b)$ to $\xi$ if and only if it converges pointwise in $(a,b)$.
to \( \xi \). This is also equivalent to \( \{\xi_k\}_{k \in \mathbb{N}} \) converges in \( C^0_{\text{loc}}(a, b) \) to \( \xi \) and \( \{\xi'_k\}_{k \in \mathbb{N}} \) converges pointwise in \((a, b)\) to \( \xi' \). Since monotone maps are derivatives of convex functions, one establishes the lemma. \( \square \)

**Lemma 6.23** Let \( \eta \in C^1_c(\mathbb{R}) \) be a nonnegative function such that \( \int_{\mathbb{R}} \eta(y) \, dy = 1 \). Suppose \( N : I \to \mathbb{R} \) is a locally bounded function and the limits \( N(x-) := \lim_{y \to x^-} N(y) \) and \( N(x+) := \lim_{y \to x^+} N(y) \) exist for all \( x \in I \). Set \( N^\varepsilon = N * \eta^\varepsilon \), where \( \eta^\varepsilon(z) = \eta(z/\varepsilon) \varepsilon^{-1} \).

Then, for any \( x \in (0,1) \), we have \( \lim_{\varepsilon \to 0} N^\varepsilon(x) = \lambda N(x+) + (1 - \lambda)N(x-) \), where \( \lambda = \int_{-\infty}^0 \eta(y) \, dy \). As a consequence, the pointwise limit exists everywhere.

**Proof.** We have \( N^\varepsilon(x) = \int_{-\infty}^0 \eta(z) N(x-z) \, dz + \int_{0}^\infty \eta(z) N(x-z) \, dz \) for every \( x \in I \). Since \( N \) is locally bounded in \( I \), the dominated convergence theorem yields the conclusion. \( \square \)

**Proposition 6.24** Suppose that \( M \in AC^q(0,T;L^p(I)) \) for some \( 1 \leq q \leq +\infty \) and \( 1 \leq p < +\infty \) and for each \( t \), the function \( M_t : I \to \mathbb{R} \) is monotone, nondecreasing and right continuous. Then, \( (t,x) \to M_t(x) \) is Borel on \( I_T \) as a function of two variables \( t \in (0,T) \) and \( x \in I \).

**Proof.** Let \( \eta \) be as in Lemma 6.23 such that \( \int_{-\infty}^0 \eta(y) \, dy = 0 \). Set \( M^k_t = M_t * \eta(1/k) \).

For \( 0 < \delta < 1/2 \), set \( I^\delta = (\delta,1-\delta) \). Then, \( \|M^k_t - M^k_s\|_{L^p(I^\delta)} \leq \|M_t - M_s\|_{L^p(I)} \) for any \( t, s \in I^\delta \). This proves that \( M^k \in AC^q(0,T;L^p(I^\delta)) \). By Lemma 6.22 we obtain the map \( (t,x) \mapsto M^k_t(x) \) is continuous on \( I^\delta_t \). By Lemma 6.23 \( \lim_{k \to +\infty} M^k_t(x) = M_t(x) \) for each \( (t,x) \in I_T \). Thus, \( M \) is Borel measurable on \( I_T \) as a pointwise limit of Borel map. \( \square \)

Observe that the spaces \( AC^q(0,T;L^p(I)) \) and \( W^{1,q}(0,T;L^p(I)) \) coincide for some \( 1 < q \leq +\infty \) and some \( 1 < p < +\infty \); according to Proposition 6.11 (see also Remark 1.1.3 [17]). If \( M \in W^{1,q}(0,T;L^p(I)) \) we denote by \( \dot{M} \in L^q(0,T;L^p(I)) \) its functional derivative and it is defined by

\[
\lim_{h \to 0} \left\| \frac{M_{T+h} - M_{T}}{h} - \dot{M}_t \right\|_{L^p(I)} = 0 \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0,T). \tag{6.23}
\]

It is straightforward to check that \( \|M'\|(t) = \|\dot{M}_t\|_{L^p(I)} \) for \( \mathcal{L}^1\text{-a.e. } t \in (0,T) \).

In the next Lemma, we shall view \( M \) as a map in \( AC^q(\mathbb{R};L^p(I)) \) by extending \( M_t = M_{0+} \) for \( t \leq 0 \) and \( M_t = M_{T-} \) for \( t \geq T \). Recall that \( \dot{M} \) can be viewed as an element of
$L^q(0, T; L^p(I))$. We obtain an extension of $\dot{M}$ to $\mathbb{R} \times I$ which we identify with an element of $L^q(\mathbb{R}; L^p(I))$.

**Lemma 6.25** Let $M \in AC^q(0, T; L^p(I))$ for some $1 < p < +\infty$, $1 < q \leq +\infty$ and let $\dot{M}$ be its functional derivative. Then

$$
\lim_{h \to 0} \int_0^T \left\| \frac{M_{t+h}-M_t}{h} - \dot{M}_t \right\|_{L^p(I)}^q \, dt = 0. \quad (6.24)
$$

As a consequence, there exist sequences $h^+_k \to 0^+, h^-_k \to 0^-$ and a measurable subset $A$ of $\mathbb{R} \times I$ such that $\mathcal{L}^2((\mathbb{R} \times I) \setminus A) = 0$ and

$$
\lim_{k \to +\infty} \frac{M_{t+h_k^+}(x) - M_t(x)}{h^+_k} = \lim_{k \to +\infty} \frac{M_{t+h_k^-}(x) - M_t(x)}{h^-_k} = \dot{M}_t(x) \quad \text{for all } (t, x) \in A. \quad (6.25)
$$

**Proof.** Set $f(t) = \|\dot{M}_t\|_{L^p(I)}$ and let $\mathcal{M} f$ be the Hardy–Littlewood maximal function given by

$$
\mathcal{M} f(t) = \sup_{0 < h < T} \left\| \int_{t-h}^{t+h} f(s) \, ds \right\|, \quad t \in \mathbb{R}. \quad (6.26)
$$

Note that, $f \in L^q(\mathbb{R})$ and so, $\mathcal{M} f \in L^q(\mathbb{R})$. Clearly, $\|(M_{t+h} - M_t)/h - \dot{M}\|_{L^p(I)} \leq \mathcal{M} f(t) + f(t)$. This, together with (6.23) and the dominated convergence theorem yields (6.24). \hfill \Box

The analogue of (5.11) in the case $p = q = 2$ is stated in the following lemma.

**Lemma 6.26** Assume that $M \in AC^q(0, T; L^p(I))$ for some $1 < q \leq +\infty$, $1 < p < +\infty$ and for each $t \in (0, T)$, the function $M_t : I \to \mathbb{R}$ is monotone nondecreasing. Let $A$ be as in the previous lemma. Suppose $(t, x), (t, \tilde{x}) \in A$ and $M_t(x) = M_t(\tilde{x})$. Then $\dot{M}_t(x) = \dot{M}_t(\tilde{x})$.

**Proof.** Without loss of generality, we can assume $M < \tilde{x}$. Then

$$
\frac{M_{t+h_k^-}(\tilde{x}) - M_{t+h_k^-}(x)}{h_k^-} \leq 0 \leq \frac{M_{t+h_k^+}(\tilde{x}) - M_{t+h_k^+}(x)}{h_k^+}
$$

By (6.25), this yields $\dot{M}_t(x) = \dot{M}_t(\tilde{x})$. \hfill \Box

Note that, the time derivative showing in the statement of Lemma 6.26 is in the sense of (6.25). Since the set $A$ defined above has full measure, we have that $\dot{M}_t(x), \dot{M}_t(\tilde{x})$ both exist in that sense for almost all $(t, x, \tilde{x}) \in (0, T) \times I \times I$. Next we show that Proposition
5.8 can be extended to any family $M$ provided that the result of the latter lemma holds. We begin with the following general result.

**Theorem 6.27** Let $N, \tilde{N} : (0, T) \times I \to \mathbb{R}$ be Borel maps. Then, $N(t,x) = N(t, \tilde{x})$ implies $\tilde{N}(t,x) = \tilde{N}(t,\bar{x})$ for $\mathcal{L}^2$-a.e. $(t,x,\bar{x}) \in (0,T) \times I^2$ if and only if there exists a Borel map $w : (0,T) \times \mathbb{R} \to \mathbb{R}$ such that

$$w(t,N(t,x)) = \tilde{N}(t,x) \quad \text{for } \mathcal{L}^2\text{-a.e. } (t,x) \in (0,T) \times I := I_T. \quad (6.27)$$

**Proof.** The direct implication is obvious. We next prove the converse in two steps.

**Step 1.** Let $\chi_I$ denote the 2-Lebesgue measure restricted to $(0, T) \times I := I_T$, $\xi : I_T \to (0, T) \times \mathbb{R}$ given by $\xi(t,x) = (t, N(t,x))$, and set $\mu := \xi_#\chi_I$. Denote by $\mu_I$ the real-valued measure whose density with respect to $\chi_I$ is $\tilde{N}$, then set $\nu := \xi_#\mu$. The components of $\nu$ are the signed measures given by $\nu(B) = \int_I \tilde{N}(t,x) d\chi_I(t,x)$ for all Borel $B \subset (0, T) \times \mathbb{R}$. Their total variations satisfy $|\nu| \ll \mu$ since $\nu(B) = \chi_I(\xi^{-1}(B))$. Thus, we can apply the Radon–Nikodým theorem for signed measures to $\nu$ and $\chi_I$. We obtain a Borel vector field $w : (0,T) \times \mathbb{R} \to \mathbb{R}$ such that $d\nu = w \, d\mu$.

**Step 2.** We now apply the disintegration theorem (see, for example, Theorem 5.3.1 [17]) to the Borel map $\xi$ and the measure $\chi_I$. Thus, for $\mu$-a.e. $(t,\bar{x}) \in (0,T) \times \mathbb{R}$, there exists a unique Borel probability measure $\chi_{t,\bar{x}}$ on $I_T$ such that the map $(t,\bar{x}) \mapsto \chi_{t,\bar{x}}(B)$ is Borel measurable for each Borel set $B \subset I_T$. Furthermore, $\chi_{t,\bar{x}}(\xi^{-1}(t,\bar{x})) = 1$ for $\mu$-a.e. $(t,\bar{x}) \in (0,T) \times \mathbb{R}$ and $\int_I f(t,x) d\chi_I(t,x) = \int_I \int_{\mathbb{R}} f(t,x) d\chi_{t,\bar{x}}(t,x) d\mu(t,\bar{x})$ for every Borel measurable $f : I_T \to [0, +\infty]$. Take $f(t,x) := \xi(t,N(t,x))\tilde{N}(t,x)$ for an arbitrary Borel map $\xi : (0,T) \times \mathbb{R} \to [0, +\infty]$. We use the previous equation to infer $\int_I \xi(t,N(t,x))\tilde{N}(t,x) d\chi_I(t,x) = \int_I \int_{\mathbb{R}} \xi(t,\bar{x}) \tilde{N}(t,x) d\chi_{t,\bar{x}}(t,x) d\mu(t,\bar{x})$. But, according to **Step 1**, we have that the integral in the left hand side above is equal to $\int_0^T \int_{\mathbb{R}} \xi(t,\bar{x}) w(t,\bar{x}) d\mu(t,\bar{x})$. The arbitrariness of $\xi$ yields $w(t,\bar{x}) = \int_{\xi^{-1}(t,\bar{x})} \tilde{N}(t,x) d\chi_{t,\bar{x}}(t,x)$ for $\mu$-a.e. $(t,\bar{x}) \in (0,T) \times \mathbb{R}$. Finally, since $\mu := \xi_#\chi_I$, we can compute $w(t,N(t,x)) = \int_{\xi^{-1}(t,N(t,x))} \tilde{N}(t,z) d\chi_{t,N(t,x)}(t,z)$ for $\chi_I$-a.e. $(t, x) \in I_T$. Note that $(t,z) \in \xi^{-1}(t,N(t,x))$ is equivalent to $N(t,z) = N(t,x)$, so we can use the hypothesis to conclude. 

The following result is a direct consequence of Theorem 6.27.
Corollary 6.28 Let $M \in W^{1,q}(0,T; L^p(I))$ for some $1 \le q \le +\infty$ and $1 \le p < +\infty$. Then, (5.11) holds if and only if there exists a Borel map $v : (0,T) \times \mathbb{R} \to \mathbb{R}$ such that
\[
v_t(M_t(x)) = \dot{M}_t(x) \quad \text{for } \mathcal{L}^2\text{-a.e. } (t,x) \in (0,T) \times I := I_T. \tag{6.28}
\]

Now we get back to (5.11). In the case of higher dimensions, A. Tudorascu [114] constructed an example showing that a statement as general as Proposition 5.8 cannot be expected to hold, even if the minimal velocity requirement on $v$ was dropped unless extra conditions are imposed. The requirement (5.11) is obviously satisfied if $M_t$ is invertible for $\mathcal{L}^1$-a.e. $t \in (0,T)$. The author of [114] showed, however, that the invertibility is, in fact, unnecessary and he gave other necessary conditions for (5.11) to hold. In the sequel, $I$ denotes the open unit interval in $\mathbb{R}$. Furthermore, for all $t \in (0,T)$ the maps $M_t$ of the family $M \in W^{1,q}(0,T; L^p(I))$ coincide $\mathcal{L}^1$-a.e. with the a.e. derivative of some monotone functions $\xi_t$ restricted to $I$. For $(t,x) \in (0,T) \times I:= I_T$ define $[M_t(x)] := \{\bar{x} \in I : M_t(x) = M_t(\bar{x})\}$. The following extra assumption
\[
\text{For almost every } (t,x) \in I_T, [M_t(x)] \text{ is either a singleton or } \mathcal{L}^1([M_t(x)]) > 0 \tag{6.29}
\]
is automatically satisfied [70]. The reason is that a “flat” portion in the graph of $M_t$ is necessarily of positive Lebesgue measure.

Proposition 6.29 Let $M \in W^{1,q}(0,T; L^p(I))$, for some $1 \le q \le +\infty$ and $1 \le p < +\infty$, be such that for $\mathcal{L}^1$-a.e. $t \in (0,T)$ the map $M_t$ coincides $\mathcal{L}^1$-a.e. with the restriction to $I$ of a derivative of a convex function defined on $\mathbb{R}$. Then, (6.29) implies (5.11).

Note that, due to the convexity of $\xi_t$ in the higher dimensions case, it is easy to see that $[M_t(x)]$ is a convex set. Thus, if it does not consist of a single element, then (6.29) implies that its interior is nonempty and convex. The author of [114] showed that (6.29) is sufficient for (5.11) to hold. The example provided in [114], however, shows a case in which neither (6.29) nor (5.11) hold.

Remark 6.30 Let $\partial_t M$ denote the distributional derivative in time of $(t,x) \to M_t(x) = M(t,x)$. Then, $\dot{M}_t(x) = \partial_t M(t,x)$ for $\mathcal{L}^2$-a.e. $(t,x) \in I_T$.
Proof. Let $\zeta \in C_c(\mathcal{C}_T)$ be arbitrary. We have

$$-\int_{I_T} \partial_t M(t,x) \zeta(t,x) \, dt \, dx = \int_{I_T} M(t,x) \partial_t \zeta(t,x) \, dt \, dx = \int_0^T dt \int_{I_T} M(t,x) \partial_t \zeta(t,x) \, dx$$

$$= \lim_{h \to 0} \int_0^T dt \int_{I_T} M(t,x) \frac{\zeta(t+h,x) - \zeta(t,x)}{h} \, dx = \lim_{h \to 0} \int dx \int_0^T M(t,x) \frac{\zeta(t+h,x) - \zeta(t,x)}{h} \, dt$$

$$= \lim_{h \to 0} \int I_T dx \int_0^T M(t,x) \frac{\zeta(t+h,x) - \zeta(t,x)}{h} \, dt$$

This concludes the proof. \qed

A consequence of Remark 6.30 is that the weak time–derivative along paths $M(\cdot, x) \in W^{1,1}(0,T)$ which is denoted by $\partial_t M(\cdot, x) \in L^1(0,T)$ coincides $L^2$–a.e. in $(0,T) \times I$ with the functional derivative $\dot{X} \in L^q(0,T; L^p(I))$ of $X$. This means we require

$$\sup_{0 < h < T} \int_0^{T-h} \left\| \frac{M_{t+h} - M_t}{h} \right\|_{L^p(I)}^q \, dt < +\infty.$$

Recall that (see Corollary 6.6), if $\rho \in AC^q(0,T; \mathcal{P}_p(\mathbb{R}))$ for some $1 \leq q \leq +\infty$ and $1 \leq p < +\infty$, then there exists at most one Borel map $(y,t) \to v_t(y) \in \mathbb{R}$ such that satisfies (6.6) and (5.5) holds in the sense of distribution. We recall $v$ the $L^1$–velocity associated to the path $\rho$. More precisely, Theorem 6.5 proves the existence and uniqueness of the $L^1$–velocity field $v : (0,T) \times \mathbb{R} \to \mathbb{R}$ such that the continuity equation is satisfied by the pair $(\rho, v)$. The natural question now is whether the $L^1$–velocity exists and satisfying

$$\int_0^T \|v_t\|_{L^p(\mu_t)}^q \, dt = \int_0^T |\rho'|^q(t) \, dt \quad \text{for } \mathcal{L}^1\text{–almost every } t \in (0,T). \quad (6.30)$$

According to Corollary 6.28, the following is true. Indeed, the proof consists of the same argument used in [114] of Proposition 3.3. The only difference is that now the optimal maps $M_t$ are not the ones pushing $\nu$ forward to $\mu_t$, but the ones pushing $\mu_0$ forward to $\mu_t$.

Proposition 6.31 Let $\rho \in AC^q(0,T; \mathcal{P}_p(\mathbb{R}))$, for some $1 \leq q \leq +\infty$ and $1 \leq p < +\infty$, be such that the family of corresponding optimal maps $M_t$ (i.e. derivatives of convex functions such that $\rho_t = M_t \# \rho_0$) satisfies condition (6.29). Then, there exists a Borel velocity map
$v : (0, T) \times \mathbb{R} \to \mathbb{R}$ such that (5.5) and (6.30) holds.

Proof. Let $T > 0$ and $\rho_0, \rho_T$ be Borel probabilities with finite $p^{th}$-moments such that Monge’s problem of optimally transporting $\rho_0$ into $\rho_T$ has a solution $M$ (see Chapters 3 and 4). Brenier [34] (see Theorem 3.25) showed that that is equivalent to the existence of a derivative $M$ of a convex function such that $M_t \# \rho_0 = \rho_1$. It is well-known, by Proposition 6.29 that, (6.29) implies (5.11). Since $M$ is invertible, we let $v_t := M_t \circ M_t^{-1}$ to see that (6.28) satisfied for $t \in (0, T)$. Consider a smooth test function $\xi(t, x)$ such that $\xi(0, \cdot) \equiv 0 \equiv \xi(T, \cdot)$. Since $M \in W^{1,q}(0, T; L^p(I))$, we have that $\int_0^T \int_I \frac{d}{dt}[\xi(t, M_t(x))] \, d\rho_0(x) \, dt = 0$ due to the boundary conditions on $\xi$. The chain rule yields $\int_0^T \int_I \{\partial_t \xi(t, M_t(x)) + M_t(x) \partial_x \xi(t, M_t(x))\} \, d\rho_0(x) \, dt = 0$, which, in view of $M_t = v_t \circ M_t$ and $\rho_t = M_t \# \rho_0$, gives the distributional form of the continuity equation $\partial_t \rho_t + \partial_x (\rho_t v_t) = 0$ from Fluid Mechanics in $(0, T) \times \mathbb{R}^d$. \hfill \Box

Notice that, the distributional form of the continuity equation $\partial_t \rho_t + \partial_x (\rho_t v_t) = 0$ can be obtained by multiply this simple identity by $\xi(t, M_t(x))$ and integrate in $x$ with respect to $\rho_0$. After that we integrate by parts in time and take into account the properties already used in the proof above.

Remark 6.32 Note that (6.29) is satisfied if all $\mu_t$ are absolutely continuous with respect to the Lebesgue measure (in which case the maps $M_t$ are invertible) or are fully supported at discrete points (convex, possibly countable, combinations of Dirac masses).

6.3.2 Optimal flow maps associated to $AC^q(0, T; \mathcal{P}_p^ac(\mathbb{R}))$ curves

We continue with the first main result of this chapter. The most important consequence of Theorem 6.16 is that, if a map $X \in W^{1,q}(0, T; L^p(I))$ solves the flow problem (5.1)–(5.2), then so does the family optimal maps $M := \{M_t\}_{t \in [0, T]}$ such that $M_\# \chi = \rho$, where $M_t$ is the monotone rearrangement of the “Lagrangian” flow maps $X_t$ associated to the absolutely continuous curves $\rho_t$ for all $t \in [0, T]$. The next theorem, which is a generalization of Proposition 5.8 in the case of higher integrability, shows the existence of the flow “optimal” map $M$ in $W^{1,q}(0, T; L^p(I))$ associated to the absolutely continuous curves with respect to Lebesgue measure $\mathcal{L}^1$. 
Theorem 6.33 Let $\rho \in AC^q(0,T; P^ac_p(\mathbb{R}))$ for some $1 \leq p < +\infty$ and some $1 \leq q \leq +\infty$. If $p = 1$, assume also that the $L^1$–velocity of $\rho$ exists. Denote by $M_t$ the optimal map pushing forward $\chi$ to $\rho_t$. Then, $M \in W^{1,q}(0,T; L^p(I))$, and it is a flow map associated to $\rho$.

Proof. Since

$$\|M(t, \cdot)\|_{L^p(I)}^p = \int_{\mathbb{R}} |y|^p \rho(t, y) \, dy = W_p^p(\rho_t, \delta_0) \leq 2^{p-1} \left( W_p^p(\rho_t, \rho_0) + \int_{\mathbb{R}} |y|^p \rho_0(y) \, dy \right) \leq 2^{p-1} \left( \left( \int_0^T \|v_t\|_{L^p(\rho_t)} \, dt \right)^p + \int_{\mathbb{R}} |y|^p \rho_0(y) \, dy \right) < +\infty \text{ for all } t \in [0,T]$$

and

$$\int_0^T \left( \int_I |v(t, M(t, x))|^p \, dx \right)^{\frac{q}{p}} dt = \int_0^T \|v_t\|_{L^q(\rho_t)}^q \, dt < +\infty,$$

it suffices to prove that $v(t, M(t, x))$ is the distributional time–derivative of $M$ to obtain the desired thesis.

Consider a standard mollifier $\eta^\varepsilon(y) = \eta(y/\varepsilon)/\varepsilon$ for $0 < \varepsilon < 1$ and let

$$\rho^\varepsilon(t, \cdot) = \eta^\varepsilon \ast \rho(t, \cdot), \quad E^\varepsilon(t, \cdot) = \eta^\varepsilon \ast [v(t, \cdot) \rho(t, \cdot)].$$

Here, $\eta \in C_c^\infty(\mathbb{R})$ is supported in $[-1,1]$, nonnegative, and even; also, $\int_{\mathbb{R}} \eta = 1$. Thus, for fixed $y \in \mathbb{R}$, $z \mapsto \eta^\varepsilon(z - y)$ is smooth and supported in $[y - \varepsilon, y + \varepsilon]$. Therefore, it can be used as a test function in (5.5) to deduce that

$$[0,T] \ni t \mapsto \int_{\mathbb{R}} \eta^\varepsilon(z - y) \rho_t(z) \, dz = \rho^\varepsilon(t, y)$$

is absolutely continuous and

$$\partial_t \rho^\varepsilon(t, y) = \int_{\mathbb{R}} \partial_z [\eta^\varepsilon(z - y)] v(t, z) \rho(t, z) \, dz$$

$$= -\int_{\mathbb{R}} (\eta^\varepsilon)'(y - z) v(t, z) \rho(t, z) \, dz = -\partial_y E^\varepsilon(t, y)$$

for a.e. $t \in [0,T]$. Now, let $F^\varepsilon(t, \cdot)$ and $F(t, \cdot)$ be the cumulative distribution functions of $\rho^\varepsilon(t, \cdot)$ and $\rho(t, \cdot)$, respectively. Note here that, since $\rho(t, \cdot) \in L^1(\mathbb{R})$, we have that
\( \rho^\varepsilon(t, \cdot) \xrightarrow{\varepsilon \to 0^+} \rho(t, \cdot) \) strongly in \( L^1(\mathbb{R}) \), which implies

\[
|F^\varepsilon(t, y) - F(t, y)| \leq \int_{-\infty}^{y} |\rho^\varepsilon(t, z) - \rho(t, z)| \, dz \xrightarrow{\varepsilon \to 0^+} 0
\]

uniformly in \( y \in \mathbb{R} \). Also, \( F^\varepsilon(t, \cdot) \) is smooth with \( \partial_y F^\varepsilon(t, y) = \rho^\varepsilon(t, y) \) for all \( t \) and \( y \).

Since \( \rho \in AC^1(0, T; \mathcal{P}_1(\mathbb{R})) \), we deduce that

\[
\int_{\mathbb{R}} |y| \rho(t, y) \, dy \leq C < +\infty \quad \text{for all } t.
\]

We also have that \( t \mapsto F^\varepsilon(t, y) \) is absolutely continuous for a.e. \( y \in \mathbb{R} \), with \( \partial_t F^\varepsilon(t, y) = -E^\varepsilon(t, y) \). Indeed, this comes as a consequence of \( \partial_t \rho^\varepsilon(t, y) = -\partial_y E^\varepsilon(t, y) \). In order to prevent some integrability issues, we also introduce a cut–off function in \( y \), namely, \( \xi_k \in C^1_c(\mathbb{R}) \) such that \( \xi_k(y) = y \) if \( |y| \leq k \), \( \xi_k(y) = 0 \) if \( |y| \geq 3k \), and

\[
|\xi_k(y)| \leq \min\{2|y|, k + 1\}, \quad |\xi'_k(y)| \leq 1
\]

for all \( y \in \mathbb{R} \). Let \( \varphi \in C^1_c(I) \) and note that

\[
[0, T] \ni t \mapsto \xi_k(y) \varphi(F^\varepsilon(t, y)) \rho^\varepsilon(t, y)
\]

is also absolutely continuous for a.e. \( y \in \mathbb{R} \), with

\[
\frac{\partial}{\partial t} [\xi_k(\varphi \circ F^\varepsilon) \rho^\varepsilon] = \xi_k(y) \varphi'(F^\varepsilon(t, y)) \partial_t F^\varepsilon(t, y) \rho^\varepsilon(t, y) + \xi_k(y) \varphi(F^\varepsilon(t, y)) \partial_t \rho^\varepsilon(t, y)
\]

\[
= -[\xi_k(y) \varphi'(F^\varepsilon(t, y)) E^\varepsilon(t, y)) \rho^\varepsilon(t, y) + \xi_k(y) \varphi(F^\varepsilon(t, y)) \partial_y E^\varepsilon(t, y)]
\]

That is, for any \( \zeta \in C^1_c(0, T) \), we have

\[
\int_0^T \zeta(t) \xi_k(y) \varphi(F^\varepsilon(t, y)) \rho^\varepsilon(t, y) \, dt = \int_0^T \zeta(t) [\xi_k(y) \varphi'(F^\varepsilon(t, y)) E^\varepsilon(t, y)) \rho^\varepsilon(t, y) + \xi_k(y) \varphi(F^\varepsilon(t, y)) \partial_y E^\varepsilon(t, y)] \, dt \quad \text{for a.e. } y \in \mathbb{R}.
\]

We would like to integrate the above equality in \( y \) over \( \mathbb{R} \), then integrate by parts over \( \mathbb{R} \) the last term in the right–hand side; for this, we need to show both sides are integrable over
Thus, \( \int_0^T \left| \frac{\partial E}{\partial t} \right| \, dt \) is independent of \( \varepsilon \) and we continue by noticing that
\[
\int_{\mathbb{R}} |y| \rho(\varepsilon(t, y)) \, dy = 
\int_{\mathbb{R}} |y| \int_{\mathbb{R}} \eta^\varepsilon(y - z) \rho(t, z) \, dz \, dy
= 
\int_{\mathbb{R}} \left( \int_{\mathbb{R}} |y| \eta^\varepsilon(y - z) \, dy \right) \rho(t, z) \, dz
\leq 
\int_{\mathbb{R}} \left( \int_{\mathbb{R}} (|y| + |y - z|) \eta^\varepsilon(y - z) \, dy \right) \rho(t, z) \, dz
= 
\int_{\mathbb{R}} |y| \rho(t, z) \, dz + \int_{\mathbb{R}} \rho(t, z) \, dz \int_{\mathbb{R}} |y| \eta^\varepsilon(y) \, dy
\leq 
\int_{\mathbb{R}} |y| \rho(t, z) \, dz + C \varepsilon, \quad \text{where } C := \int_{-1}^1 |y| \eta(y) \, dy.
\]
Thus, \( \int_{\mathbb{R}} |\xi_k(y)| \rho(\varepsilon(t, y)) \, dy \) is bounded by a finite constant which is independent of \( t \) and \( 0 < \varepsilon < 1 \). Thus, \( (t, y) \mapsto \xi_k(y) \varphi(F^\varepsilon(t, y)) \rho^\varepsilon(t, y) \) is in \( L^\infty(0, T; L^1(\mathbb{R})) \), with bounds independent of \( \varepsilon \in (0, 1) \) and \( k \). As for the right-hand side of (6.31), we see that
\[
|E^\varepsilon(t, y)| \leq \int_{\mathbb{R}} \eta^\varepsilon(y - z) |v(t, z)| \rho(t, z) \, dz \leq \frac{1}{\varepsilon} \max_{\mathbb{R}} |\eta||v_t||L^1(\rho_t)|.
\]
Thus,
\[
\int_{\mathbb{R}} \left| \xi_k(y) \varphi'(F^\varepsilon(t, y)) E^\varepsilon(t, y) \rho^\varepsilon(t, y) \right| \, dy \leq \frac{C}{\varepsilon} ||v_t||_{L^1(\rho_t)} ||\varphi'||_{\infty},
\]
where we absorbed the uniform bound on \( \int_{\mathbb{R}} |\xi_k(y)| \rho^\varepsilon(t, y) \, dy \) proved above in the constant \( C \). Note that the right hand side of this inequality lies in \( L^1(0, T) \), so
\[
(t, y) \mapsto \xi_k(y) \varphi'(F^\varepsilon(t, y)) E^\varepsilon(t, y) \rho^\varepsilon(t, y)
\]
is in \( L^1((0, T) \times \mathbb{R}) \) (even though, in this case, the bound may be of order \( \varepsilon^{-1} \)).

Finally, the last term in (6.31) is \( \xi_k(y) \varphi(F^\varepsilon(t, y)) \partial_y E^\varepsilon(t, y) \), and it satisfies
\[
\int_{\mathbb{R}} \left| \xi_k(y) \varphi(F^\varepsilon(t, y)) \partial_y E^\varepsilon(t, y) \right| \, dy \leq (k + 1)||\varphi||_{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}} |(\eta^\varepsilon)'(y - z)| |v(t, z)| \rho(t, z) \, dz \, dy
= (k + 1)||\varphi||_{\infty} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} |(\eta^\varepsilon)'(y - z)| \, dy \right) |v(t, z)| \rho(t, z) \, dz
= \varepsilon^{-1}(k + 1)||\varphi||_{\infty} ||\eta'||_{L^1(\mathbb{R})} ||v(t, \cdot)||_{L^1(\rho(\cdot))},
\]
so it lies in $L^1((0, T) \times \mathbb{R})$ as well. Since $[0, T] \ni t \mapsto \|v(t, \cdot)\|_{L^1(\rho(t, \cdot))}$ lies in $L^1(0, T)$, we deduce that (6.31) can be integrated with respect to $y$ over $\mathbb{R}$. Then, after a spatial integration by parts of the last term in the right–hand side (which leads to the cancelation of the first term in the right–hand side), we may apply Fubini’s theorem to discover that

$$
\int_0^T \dot{\zeta}(t) \int_{\mathbb{R}} \xi_k(y) \varphi(F^\varepsilon(t, y)) \rho^\varepsilon(t, y) dy dt = - \int_0^T \zeta(t) \int_{\mathbb{R}} \xi_k'(y) \varphi(F^\varepsilon(t, y)) E^\varepsilon(t, y) dy dt,
$$

(6.32)

with integrands in $L^1((0, T) \times \mathbb{R})$.

Next, we let $\varepsilon \to 0^+$ and use the uniform convergence of $F^\varepsilon(t, \cdot)$ to $F(t, \cdot)$ and the $L^1(\mathbb{R})$–convergence of $\rho^\varepsilon(t, \cdot)$ to $\rho(t, \cdot)$, along with that of $E^\varepsilon(t, \cdot)$, to $v(t, \cdot)\rho(t, \cdot)$ to infer that for each $t \in [0, T]$ we have (since $\xi_k$ is continuous and compactly supported)

$$
U^\varepsilon(t) := \int_{\mathbb{R}} \xi_k(y) \varphi(F^\varepsilon(t, y)) \rho^\varepsilon(t, y) dy \xrightarrow{\varepsilon \to 0^+} \int_{\mathbb{R}} \xi_k(y) \varphi(F(t, y)) \rho(t, y) dy
$$

and

$$
V^\varepsilon(t) := \int_{\mathbb{R}} \xi_k'(y) \varphi(F^\varepsilon(t, y)) E^\varepsilon(t, y) dy \xrightarrow{\varepsilon \to 0^+} \int_{\mathbb{R}} \xi_k'(y) \varphi(F(t, y)) v(t, y) \rho(t, y) dy.
$$

By some well–known convolution properties of $L^1$–functions (note also that $\rho(t, \cdot)$ is nonnegative, as a probability density), we have

$$
|U^\varepsilon(t)| \leq (k + 1) \|\varphi\|_{\infty} \quad \text{and} \quad |V^\varepsilon(t)| \leq (k + 1) \|\varphi\|_{\infty} \|v(t, \cdot)\|_{L^1(\rho(t, \cdot))}.
$$

Next we let $\varepsilon \to 0^+$ in (6.32) and use Dominated Convergence over $[0, T]$ to get

$$
\int_0^T \dot{\zeta}(t) \int_{\mathbb{R}} \xi_k(y) \varphi(F(t, y)) \rho(t, y) dy dt = - \int_0^T \zeta(t) \int_{\mathbb{R}} \xi_k'(y) \varphi(F(t, y)) v(t, y) \rho(t, y) dy dt.
$$

But $|\xi_k(y)| \leq 2|y|$, $\|\xi_k'\|_{\infty} \leq 1$ and $\xi_k(y)$ and $\xi_k'(y)$ converge pointwise to $y$ and 1, respectively, for all $y \in \mathbb{R}$. Recall that the first moment of $\rho(t, \cdot)$ is bounded uniformly with respect to $t$, and $v\rho \in L^1((0, T) \times \mathbb{R})$. Thus, we may let $k \to \infty$ and use Dominated Convergence
Theorem on \([0,T] \times \mathbb{R}\) to get
\[
\int_0^T \dot{\zeta}(t) \int_{\mathbb{R}} y \varphi(F(t,y)) \rho(t,y) \, dy \, dt = - \int_0^T \zeta(t) \int_{\mathbb{R}} \varphi(F(t,y)) v(t,y) \rho(t,y) \, dy \, dt. \tag{6.33}
\]
We now use the fact that (see Remark 6.18)
\[
\rho(t,\cdot) \ll L^1 \implies F(t,M(t,x)) = x \text{ for a.e. } x \in (0,1), \tag{6.34}
\]
and also the fact that \(M_t \# \chi = \rho_t\), to conclude
\[
\int_0^T \dot{\zeta}(t) \int_0^1 \varphi(x) M(t,x) \, dx \, dt = - \int_0^T \zeta(t) \int_0^1 v(t,M(t,x)) \varphi(x) \, dx \, dt
\]
for all \(\zeta \in C^1_c(0,T), \varphi \in C^1_c(I)\). Thus, the distributional time–derivative of \(M(t,x)\) is \(v(t,M(t,x))\). Of course, the last displayed equality and the uniform \(L^q - L^p\) bounds obtained in the first paragraph of this proof also imply that for a.e. \(x \in I\) the function \(t \mapsto M(t,x)\) is absolutely continuous on \([0,T]\) and its a.e. time derivative is \(v(\cdot,M(\cdot,x))\). Thus, \(M\) is as in Definition 5.1.

Note that, Theorem 6.33 shows that a curve \(\rho \in AC^q(0,T;\mathcal{P}_p(\mathbb{R}))\) whose range is in \(\mathcal{P}_p^{ac}(\mathbb{R})\) and admits \(L^1\)–velocity (which is not always the case if \(p = 1\)) generates a Borel flow by the monotone rearrangement of \(\rho\).

Now, let \(X_0 : I \to \mathbb{R}\) such that \(X_{0\#}\chi = \rho_0 \ll L^1\). If \(F_0\) is the cumulative distribution function of \(\rho_0\), and \(M_0 : I \to \mathbb{R}\) is the optimal map pushing \(\chi\) forward to \(\rho_0\), then \(g_0 := F_0 \circ X_0\) is the Lebesgue a.e. unique map such that \(g_{0\#}\chi = \chi\) and \(X_0 = M_0 \circ g_0\). This is the polar decomposition of a map \(X_0\), due to Theorem 3.25 (for more details we refer to Subsection 3.3.3 that presents the original approach of polar factorization by Brenier [34]).

**Corollary 6.34** Let \(\rho \in AC^q(0,T;\mathcal{P}_p^{ac}(\mathbb{R}))\) for some \(1 \leq p < +\infty\) and some \(1 \leq q \leq +\infty\). If \(p = 1\), assume also that the \(L^1\)–velocity of \(\rho\) exists. Then, for any \(X_0 : I \to \mathbb{R}\) such that \(X_{0\#}\chi = \rho_0\), there exists a flow map \(X \in W^{1,q}(0,T;L^p(I))\) associated to \(\rho\) that starts at \(X_0\). More precisely, \(X\) can be chosen such that one has \(X_t = M_t \circ g_0\) for all \(t \in [0,T]\), where \(M\) is the family of optimal maps associated to \(\rho\), and \(g_0\) is the measure–preserving map such that \(X_0 = M_0 \circ g_0\).
Proof. Let $X_0 = M_0 \circ g_0$ be the polar decomposition of $X_0$ as recalled above. There exists a Borel set $A \subset I$ such that $\chi(A) = 1$ and

$$M(t, z) = M_0(z) + \int_0^t v(s, M(s, z))ds \quad \text{for all } t \in [0, T] \text{ and all } z \in A.$$ 

But $1 = \chi(A) = \chi(g_0^{-1}(A))$ due to $g_0#\chi = \chi$, so $g_0(x) \in A$ for $L^1$–a.e. $x \in I$. Thus,

$$M(t, g_0(x)) = M_0(g_0(x)) + \int_0^t v(s, M(s, g_0(x)))ds \quad \text{for all } t \in [0, T] \text{ and } L^1$–a.e. $x \in A.$

Thus, $X_t := M_t \circ g_0$ satisfies the flow problem (5.1)--(5.2) with $X(0, \cdot) \equiv X_0$. \qed

6.3.3 Optimal flow maps associated to $AC^q(0, T; \mathcal{P}_p(\mathbb{R}))$ curves

The discussion in the preceding subsection can be continued in the case where $\rho$ does not necessarily consist of probabilities that are absolutely continuous with respect to the Lebesgue measure. For example, an interesting question is whether any $\rho \in AC^1(0, T; \mathcal{P}_p(\mathbb{R}))$ for some $p > 1$ (in which case $v$ exists according to Ambrosio et al. [17]) admits a Lagrangian description as a flow consisting of the optimal maps $M_t$ such that $M_t#\chi = \rho_t$. Note that the assumption $\rho_t \ll L^1$ for all $t \in [0, T]$ was essential to the proof of Theorem 6.33, (and, consequently, of Corollary 6.34 as well). This question was settled in the affirmative if $q > 1$ by the next theorem, but $q = 1$ remained open.

In case $1 < p < +\infty$ and $1 < q \leq +\infty$, we have another proof different than the one of Theorem 6.33 above, for which the absolute continuity of $\rho(t, \cdot)$ with respect to the Lebesgue measure is unnecessary. Also, note that $p > 1$ also guarantees the existence [17] of the $L^1$–velocity for the curve $\rho$ considered below.

**Theorem 6.35** Let $\rho \in AC^q(0, T; \mathcal{P}_p(\mathbb{R}))$ for some $1 < p < +\infty$ and some $1 < q \leq +\infty$, and let $M$ be the corresponding family of optimal maps such that $M_t#\chi = \rho_t$. Then, $M \in W^{1,q}(0, T; L^p(I))$, and it is a flow map associated to $\rho$ and starting at $M_0$.

**Proof.** We have $W_p(\rho_s, \rho_t) = \|M_s - M_t\|_{L^p(I)}$, so there exists $\beta \in L^q(0, T)$ (can take
\[ \beta(t) := \|v_t\|_{L^p(\rho_t)} \]
such that
\[ \|M_s - M_t\|_{L^p(I)} \leq \int_s^t \beta(\tau) \, d\tau \quad \text{for all } 0 \leq s \leq t \leq T. \]

Fix \( t \in (0, T) \)–a Lebesgue point for \( \beta \), and let \( h \) be sufficiently small such that \( t + h \in (0, T) \). Then,
\[ \left\| \frac{M_{t+h} - M_t}{h} \right\|_{L^p(I)} \leq \int_t^{t+h} \beta(s) \, ds \leq M\beta(t), \]
where \( M\beta \) is the Hardy–Littlewood maximal function associated to \( \beta \in L^q(0, T) \) given by (6.26). Since \( q > 1 \), we have \( M\beta \in L^q(0, T) \) and \( \|M\beta\|_{L^q(0, T)} \leq C \|\beta\|_{L^q(0, T)} \). Thus, by setting \( \beta(t) = \|v_t\|_{L^p(\rho_t)} \) we have
\[ \int_0^T \left\| \frac{M_{t+h} - M_t}{h} \right\|_{L^p(I)}^q \, dt \leq C(q) \int_0^T \|v_t\|_{L^p(\rho_t)}^q \, dt < +\infty. \]

But \( L^q(0, T; L^p(I)) \) is a reflexive Banach space for \( 1 < p < +\infty \) and \( 1 < q < +\infty \) (see, e.g., [77]), so there exists \( u \in L^q(0, T; L^p(I)) \) and \( h_n \to 0 \) such that
\[ \frac{M(\cdot + h_n, \cdot) - M}{h_n} \to u \quad \text{in the sense of distributions on } (0, T) \times I. \]

The same holds if \( q = +\infty \) (as \( L^\infty(0, T; L^p(I)) \subset L^q(0, T; L^p(I)) \)). Thus,
\[ - \int_0^T \int_I \phi(t, x) M(t, x) \, dx \, dt = \lim_{n \to \infty} \int_0^T \int_I \phi(t, x) \frac{M(t + h_n, x) - M(t, x)}{h_n} \, dx \, dt \]
\[ = \int_0^T \int_I \phi(t, x) u(t, x) \, dx \, dt \quad \text{for all } \phi \in C_c^\infty((0, T) \times I) \]
and it follows that \( M \in W^{1,q}(0, T; L^p(I)) \) with functional derivative \( u = :M \in L^q(0, T; L^p(I)) \).

In particular, for a.e. \( x \in I, \ t \mapsto M(t, x) \) is absolutely continuous and \( \partial_t M(t, x) = \dot{M}(t, x) \)
for a.e. \( t \in (0, T) \). Thus,
\[ M(t + h, x) - M(t, x) = \int_t^{t+h} \dot{M}(s, x) \, ds \quad \text{for all } t \in (0, T) \]
and \( h \in \mathbb{R} \) such that \( 0 \leq t + h \leq T \) implies
\[
\left\| \frac{M_{t+h} - M_t}{h} - \dot{M}(t, \cdot) \right\|_{L^p(I)} \leq \int_t^{t+h} \left\| \dot{M}(s, \cdot) - \dot{M}(t, \cdot) \right\|_{L^p(I)} \, ds \quad \text{for all } t \in (0, T)
\]
which converges to 0 for a.e. \( t \in (0, T) \) (almost all \( t \in (0, T) \) are Lebesgue points for the Bochner integral of \( t \mapsto \dot{M}(t, \cdot) \in L^q(0, T; L^p(I)) \)). Furthermore, since
\[
\int_t^{t+h} \left\| \dot{M}(s, \cdot) - \dot{M}(t, \cdot) \right\|_{L^p(I)} \, ds \leq f(t) + \mathcal{M}f(t) \in L^q(0, T),
\]
where \( f(t) := \|\dot{M}(t, \cdot)\|_{L^p(I)} \) and \( \mathcal{M}f \) is its Hardy–Littlewood maximal function, we infer that
\[
\int_0^T \left\| \frac{M_{t+h} - M_t}{h} - \dot{M}_t \right\|_{L^p(I)}^q \, dt \xrightarrow{h \to 0} 0.
\]
As in Lemma 6.25, this implies the existence of a set \( A \subset (0, T) \times I \) of full Lebesgue measure and of sequences \( h_n^+ \to 0^+ \) such that
\[
\lim_{n \to \infty} \frac{M(t + h_n^+, x) - M(t, x)}{h_n^+} = \dot{M}(t, x) \quad \text{for all } (t, x) \in A. \tag{6.35}
\]
Since \( [0, T] \ni t \mapsto \int_{\mathbb{R}} \varphi(y) \rho_t(dy) \) is absolutely continuous for all \( \varphi \in C_c^\infty(\mathbb{R}) \) and \( \int_{\mathbb{R}} \varphi(y) \rho_t(dy) = \int_I \varphi(M(t, x)) \, dx \), (6.35) easily implies that
\[
\frac{d}{dt} \int_{\mathbb{R}} \varphi(y) \rho_t(dy) = \int_I \dot{M}(t, x) \varphi'(M(t, x)) \, dx.
\]
Also, in Corollary 6.28 it is proved that, there exists \( w : (0, T) \times \mathbb{R} \to \mathbb{R} \) a Borel map such that \( \dot{M}(t, x) = w(t, M(t, x)) \) for a.e. \( (t, x) \). Therefore,
\[
\frac{d}{dt} \int_{\mathbb{R}} \varphi(y) \rho_t(dy) = \int_{\mathbb{R}} \varphi'(y) w(t, y) \rho_t(dy) \, dx \quad \text{for all } \varphi \in C_c^\infty(\mathbb{R}),
\]
which, since \( w_t \in L^p(\rho_t) \) for a.e. \( t \) and \( t \mapsto \|w_t\|_{L^p(\rho_t)} \in L^q(0, T) \), we deduce that (by the uniqueness of the velocity associated to \( t \mapsto \rho_t \) — see Theorem 6.5) for a.e. \( t \in (0, T) \), \( v_t \equiv w_t \) in the \( \rho_t \)-a.e. sense. \( \square \)

**Conclusion 6.36** In all points below, the connection between \( \rho \) and \( M \) is that \( M_t \# \chi = \rho_t \).
(1) We have showed in Theorem 6.33 that if \( \rho \in AC^q(0,T;\mathcal{P}_p(\mathbb{R})) \) and its \( L^1 \)-velocity exists, then \( M \in W^{1,q}(0,T;L^p(I)) \) for any \( 1 \leq p < +\infty \) and \( 1 \leq q \leq +\infty \). It is not clear how to obtain (6.33) (which is essential to the proof of Theorem 6.33) if we drop the absolute continuity of \( \rho(t,\cdot) \) with respect to \( \mathcal{L}^1 \). Note also that if \( p = 1 \) said theorem surmises the existences of the \( L^1 \)-velocity \( v \), as its existence is unclear in this case. For \( 1 < p < +\infty \) we know from [17] that \( v \) exists.

(2) If we restrict ourselves to \( 1 < p < +\infty \) and \( 1 < q \leq +\infty \), then the \( L^1 \)-velocity exists and we can do away with the restriction \( \rho_t \in \mathcal{P}_{ac}^p(\mathbb{R}) \) to conclude that \( \rho \in AC^q(0,T;\mathcal{P}_p(\mathbb{R})) \) implies \( M \in W^{1,q}(0,T;L^p(I)) \). In both cases (i.e. (1) and (2)), \( M \) is a flow map for the velocity \( v \) associated to the path \( \rho \).

(3) In fact, combining Theorem 6.16 and Theorem 6.35 we deduce the powerful fact that if \( 1 < p < +\infty \) and \( 1 < q \leq +\infty \), then \( M \in W^{1,q}(0,T;L^p(I)) \) is equivalent to \( \rho \in AC^q(0,T;\mathcal{P}_p(\mathbb{R})) \).

Note that, we do not know if every \( \rho \in AC^1(0,T;\mathcal{P}_1(\mathbb{R})) \) admits an \( L^1 \)-velocity field. Furthermore, in connection to (1) of the conclusion above, even if the velocity exists, it is still unknown whether \( M \) will be a flow map for it. These questions will be investigated in future work.

Finally, we would like to make it clear that, we had not stated explicitly that we were only concerned with \( AC^q(0,T;\mathcal{P}_p(\mathbb{R})) \) curves for which the velocity \( v \) existed; after all, our main goal is to study uniqueness for the flow equation (5.1)−(5.2), which is regarded as the Lagrangian description of a probability distribution on the line which evolves driven by a velocity \( v \) according to the continuity equation (5.5) (Eulerian description). Thus, a physical velocity driving the flow \( X \) is necessary to make sense of both descriptions (Lagrangian and Eulerian) and their connection \( \mu_t = X_{t\#}\chi \). So, we tacitly assumed that all probability curves \( \mu_t \) considered admitted a velocity.
Chapter 7

Uniqueness for the Lagrangian flow and for the Continuity Equation

The aim of this Chapter is to develop a well-posedness theory for Lagrangian flows in which the usual assumptions of the velocity as in [6], [56] is replaced by an absolute continuity of probability curves (or, densities). This is particularly important in view of the applications, for instance to the uniqueness of a solution of the continuity equation (see Corollary 7.10). A uniqueness treatment of this topic is done in the book of Ambrosio, Gigli, and Savaré [17]. Since the absolute continuity assumption on the density is necessary in order to give a distributional meaning to the continuity equation (see Theorem 8.3.1 of [17]), we deal here with the probability density in continuity equation form. We will consider again the case of smooth vector fields, from the ODE viewpoint, in Section 5.1. In this Chapter, we address the main motivation behind the thesis: conditions under which the a Borel “Lagrangian” flow map is unique (in particular, monotone nondecreasing), given the velocity vector field \( \mathbf{v} \in L^q(0,T;L^p(\mathbb{R})) \) corresponding to the density \( \rho \in AC^q(0,T;\mathcal{P}_p(\mathbb{R})) \) for some \( 1 \leq p < +\infty \) and some \( 1 \leq q \leq +\infty \). Due to the absolute continuity curve of probability measures on the real line, which admits a Lagrangian description as a well-defined flow of its velocity field, the assumptions needed for the uniqueness are completely different than those of Chapter 1. We are able to show two different kinds of some sufficient assumptions on such curve and/or its velocity that ensure that the only flow map associated with the curve consists of a time independent rearrangement of the generalized inverses of...
the cumulative distribution functions of the measures on the curve.

We start in the first section by presenting some standard considerations and collecting some results available in the literature; we also mention a first main result (Theorem 7.5) obtained with A. Tudorascu [21] by a direct method (loosely connected to the narrative from Chapter 5), which goes beyond the time–space joint continuity assumption on \( \rho \) and \( v \). Subsection 7.1.1 is devoted to provide some regularity of the cumulative distribution function as a one of the main tools pertaining to the present approach. In Subsection 7.1.2, we state and prove a first main theorem in this thesis regarding the uniqueness of Lagrangian description of jointly continuous of a \( AC^1(0, T; P_{ac}^1(\mathbb{R})) \) curve. Subsection 7.1.3 is devoted to the \( L^1 \)–velocity of jointly continuous of a \( AC^1(0, T; P_{ac}^1(\mathbb{R})) \) curve. In Subsection 7.1.4, we present the uniqueness of solutions for the continuity equation. The second section is devoted to the presentation of a work also in collaboration with A. Tudorascu [21], in which sharp well–posedness result in the higher integrability case are obtained. As a byproduct of our approach, uniqueness for the continuity equation (1.6) also arises within a precisely defined class of solutions; within that class, the velocity \( v \) will be a single curve velocity. In Subsection 7.2.1 we introduce the functional derivative of optimal maps in the integrable space \( L^q(0, T; L^p(I)) \) for some \( 1 \leq p < +\infty \) and some \( 1 < q \leq +\infty \). In Subsection 7.2.2, we present the uniqueness of the Lagrangian flows within the larger class of time–Sobolev spaces \( W^{1,1}(0, T; L^1(\mathbb{R})) \). We present more deeply indebted to the results from Chapter 5 as it analyzes the case where the densities enjoy some precisely quantified integrability (no conditions imposed on the velocity in this case). In Theorem 7.14, which is the second main result of this thesis, we provide uniqueness under some integrability assumption for \( \rho \), without assuming continuity of either \( \rho \) nor \( v \).

### 7.1 Continuous Case

This section demonstrates the uniqueness of Lagrangian description under the assumption \( \rho, v \in C([0, T] \times \mathbb{R}) \) and \( \rho \in AC^1(0, T; P_1(\mathbb{R})) \) with \( \rho \ll \mathcal{L}^1 \). These regularities have some new and interesting corollaries, regarding uniqueness for the continuity equation (1.6). The uniqueness of the continuity equation holds under additional assumptions: we must require
the sufficient conditions on the velocity to render it a single curve velocity with respect to \( \rho \in AC^1(0, T; P^{ac}_1(\mathbb{R})) \) (see Proposition 7.8 and Corollary 7.10), which turns out to be not necessary, in view of a counterexample (Example 7.11 below). These conditions allow to prove a jointly continuous in time–space of a \( AC^1(0, T; P^{ac}_1(\mathbb{R})) \) curve for solutions to the continuity equation in spite of \( v \) being much less regular. In particular, there exists a unique solution to the continuity equation for given initial \( \rho_0 \in P^{ac}_1(\mathbb{R}) \), as well as the corresponding Borel flow.

### 7.1.1 Regularity of the cumulative distribution function

We start this subsection with the following elementary lemma, which recalls one of the standard properties for the space of continuously differentiable functions with compact support in \( \mathbb{R} \).

**Lemma 7.1** If \( I \subset \mathbb{R} \) is an open interval, then \( C^1_c(I) \) is separable with respect to the \( C^1(I) \) topology; that is, there exists a sequence \( \{\xi_n\}_{n \geq 1} \subset C^1_c(I) \) such that it is dense in \( C^1_c(I) \) and \( \{\xi_n'\}_{n \geq 1} \) is dense in \( \{\xi' : \xi \in C^1_c(I)\} \) (with respect to uniform convergence).

**Proof.** Assume first that \( I = (a, b) \) is bounded, and also fix \( \varphi \in C_c(I) \) with \( \bar{\varphi} := \int_I \varphi = 1 \). Define the operator \( S : C_c(I) \rightarrow C^1_c(I) \) by

\[
S\xi(x) = \int_a^x \xi(z) \, dz - \bar{\xi} \int_a^x \varphi(z) \, dz
\]

(note that, indeed, this acts between the specified spaces). We next see that \( S \) is onto by checking that \( S(f') = f \) for any \( f \in C^1_c(I) \) (see Remark 7.2 below). It is also linear and continuous with respect to the sup norm. Indeed, since we have \( (S\xi)'(x) = \xi(x) - \bar{\xi} \varphi(x) \), we also have

\[
\|S\xi\|_{\infty} + \|(S\xi)'\|_{\infty} \leq C \|\xi\|_{\infty}
\]

for some \( C < \infty \) independent of \( \xi \in C_c(I) \).

Thus, if \( \{\xi_n\}_n \subset C_c(I) \) is dense in \( C_c(I) \) with respect to the sup norm, we deduce that \( \{S\xi_n\}_n \) is dense in \( C^1_c(I) \) with respect to the \( C^1(I) \) norm; that is, the separability of \( C_c(I) \)
with the sup norm implies the separability of $C^1_c(I)$ with respect to the $C^1_c(\bar{I})$–norm. Finally, if $I$ is unbounded, we can write it as a countable union of bounded intervals to conclude. □

**Remark 7.2** To check that $S$ is onto, take $f \in C^1_c(a, b)$ and refer to $\int_0^\xi(z) \, dz - \bar{\xi} \int_0^\varphi(z) \, dz = f(x)$. Differentiate both sides to obtain $\xi(x) - \bar{\xi} \varphi(x) = f'(x)$. Therefore, we have $\xi(x) = f'(x) + \bar{\xi} \varphi(x)$, or

$$\xi(x) = f'(x) + c \varphi(x), \quad \text{where } c \in \mathbb{R}. \quad (7.1)$$

Indeed, since $\int_a^b f'(x) \, dx = 0$ and $\int_I \varphi = 1$, we have that $(7.1)$ implies $c = \bar{\xi}$. Thus, we have $S \xi(x) = \int_0^x [f'(z) + c \varphi(z)] \, dz - c \int_0^x \varphi(z) \, dz = f(x)$, as required.

To prove a chain rule for cumulative distribution functions (see Remark 7.4), we need the following fundamental statement, which provides a weak time–differentiability of the cumulative distribution functions.

**Proposition 7.3** Let $I, J$ be two open intervals, $J$ be unbounded below and $U = (\rho, w) \in L^1(I \times J; \mathbb{R}^2)$ with $\text{div } U = 0$ in the sense of distributions. Then, for almost every $y \in J$ the function $F(t, y) = \int_{-\infty}^y \rho(t, z) \, dz$ lies in $W^{1,1}(I)$ and $\dot{F}(\cdot, y) = -w(\cdot, y)$. Furthermore, if $J \ni y \mapsto w(\cdot, y)$ is continuous in the space endowed with weak topology (i.e., in $L^1_{\text{loc}}(I)$–weak), and $\rho(t, \cdot) \in L^1(I)$ for all $t \in I$, then the above conclusions hold for all $y \in J$.

**Proof.** For every $\xi \in C^1_c(I)$ and $\zeta \in C^1_c(J)$, we have

$$\int_I \dot{\xi}(t) \left( \int_J \rho(t, z) \zeta(z) \, dz \right) \, dt = -\int_I \xi(t) \left( \int_J w(t, z) \zeta'(z) \, dz \right) \, dt. \quad (7.2)$$

Fix $k \in \mathbb{Z}, k \geq 2$, and consider:

$$\zeta_k(z) = \begin{cases} 
0 & \text{if } z \in (-\infty, -k - 1], \\
z + k + 1 & \text{if } z \in (-k - 1, -k], \\
1 & \text{if } z \in (-k, y], \\
-kz + ky + 1 & \text{if } z \in (y, y + 1/k], \\
0 & \text{if } z > y + 1/k, 
\end{cases}$$
which is continuous, compactly supported and piecewise linear on \( \mathbb{R} \). Let \( \{ \zeta^n \}_{n \geq 1} \subset C^1_c(J) \) be such that \( \zeta^n \to \zeta_k \) uniformly, and

\[
(\zeta^n)' \to \zeta_k' \quad \text{everywhere except at } z = -k - 1, -k, y, y + \frac{1}{k}
\]

(where \( \zeta_k \) is not differentiable there) and such that \( \| (\zeta^n)' \|_{\infty} \leq 2k \). Then, we can pass to the limit in (7.2) with \( \zeta \equiv \zeta^n \) to get

\[
\int_I \dot{\xi}(t) \int_J \rho(t, z) \zeta_k(z) \, dz \, dt = -\int_I \xi(t) \left[ \int_{-k-1}^{-k} w(t, z) \, dz - k \int_y^{y+\frac{1}{k}} w(t, z) \, dz \right] \, dt.
\]

Since \( \rho, w \in L^1(I \times J) \), we can pass to the limit in

\[
\int_I \left[ \dot{\xi}(t) \int_J \rho(t, z) \zeta_k(z) \, dz + \xi(t) \int_{-k-1}^{-k} w(t, z) \, dz \right] \, dt
\]

by the Dominated Convergence Theorem to get (as \( k \to \infty \)):

\[
\lim_{k \to \infty} \int_I \int_y^{y+\frac{1}{k}} w(t, z) \xi(t) \, dz \, dt = \int_I \dot{\xi}(t) F(t, y) \, dt \tag{7.3}
\]

By Lemma 7.1, there exists a sequence \( \{ \xi_n \}_{n \geq 1} \subset C^1_c(I) \) dense in \( C^1_c(I) \) with respect to the \( C^1(\bar{I}) \) topology. Fix such \( \xi \equiv \xi_n \) in (7.3) above. The fact that

\[
z \mapsto \int_I w(t, z) \xi_n(t) \, dt \in L^1(J)
\]

implies there exists a sequence of measurable subsets \( A_n \) of \( J \) with \( \mathcal{L}^1(J \setminus A_n) = 0 \) such that every \( y \in A_n \) is a Lebesgue point for this mapping. Thus, at any such \( y \) we have

\[
\lim_{k \to \infty} \int_I \int_y^{y+\frac{1}{k}} w(t, z) \xi_n(t) \, dz \, dt = \lim_{k \to \infty} \int_y^{y+\frac{1}{k}} \int_I w(t, z) \xi_n(t) \, dt \, dz = \int_I \xi_n(t) w(t, y) \, dt. \tag{7.4}
\]

Let \( A = \bigcap_n A_n \), so that \( \mathcal{L}^1(J \setminus A) = 0 \) and (7.4) holds for all \( y \in A \) and all \( n \geq 1 \). By (7.3), we get

\[
\int_I \xi_n(t) w(t, y) \, dt = \int_I \dot{\xi}_n(t) F(t, y) \, dt \quad \text{for all } n \geq 1 \text{ and all } y \in A.
\]

By the density described above and the fact that \( F(\cdot, y) \in L^1(I) \) and \( w(\cdot, y) \in L^1(J) \) for
almost every \( y \in J \), we deduce
\[
\int_I \dot{\xi}(t) F(t, y) \, dt = \int_I \xi(t) w(t, y) \, dt \quad \text{for a.e. } y \in J \text{ and all } \xi \in C_c^1(0, T).
\]

Thus, we conclude that the function \( t \mapsto F(t, y) \in W^{1,1}(I) \) for almost every \( y \in J \), and its
distributional derivative is \( \dot{F}(\cdot, y) = -w(\cdot, y) \). In particular,
\[
F(b, y) - F(a, y) = -\int_a^b w(t, y) \, dt \tag{7.5}
\]
for all \( a, b \in I \) with \( a \leq b \) and almost every \( y \in J \), and we will use this to prove the
second statement. Pick an arbitrary \( y_0 \in J \) and consider a sequence \( \{y_n\}_{n \geq 1} \subset J \) such that
\( y_n \to y_0 \) and (7.5) holds for \( y = y_n \) for all \( n \geq 1 \). Then, we get that (7.5) holds for \( y_0 \) as
well by passing to the limit as \( n \to \infty \). Indeed, \( \rho(t, \cdot) \in L^1(J) \) for all \( t \in I \) implies \( F(t, \cdot) \) is
continuous on \( J \) for all \( t \in I \). To pass to the limit in the right–hand side we use that \( w(\cdot, y_n) \)
converges to \( w(\cdot, y_0) \) weakly in \( L^1(a,b) \). Thus, (7.5) holds for all \( a, b \in I \) and all \( y \in J \).

**Remark 7.4** Since \( \rho(t, \cdot) \in L^1(J) \) for almost every \( t \in I \), we infer \( F(t, \cdot) \in L^\infty(J) \) with
spatial derivative \( \partial_y F(t, \cdot) = \rho(t, \cdot) \in L^1(J) \). If \( w = v \rho \) for some Borel map \( v = v(t, y) \) and
\( \dot{X}(t, x) = v(t, X(t, x)) \) in some well–defined sense, then a formal calculation reveals
\[
\begin{align*}
\partial_t[F(t, X(t, x))] &= \dot{F}(t, X(t, x)) + \partial_y F(t, X(t, x)) \dot{X}(t, x) \\
&= -w(t, X(t, x)) + \rho(t, X(t, x)) v(t, X(t, x)) = 0. \tag{7.6}
\end{align*}
\]

This means that the solution \( F \) is constant on the characteristic lines of \( v \). Thus, provided
that \( \rho(0, \cdot) = \chi \) and \( X(0, \cdot) = \text{Id}_I \), we deduce \( F(t, X(t, x)) = x \) for all \( t \in [0, T] \) and
almost every \( x \in I \), to be a solution of the transport equation (5.6) with an initial data
\( F(0, x) = F_0(x) \). This fact has far reaching consequences, as we shall see below.

Note that, (7.6) is equivalent to saying that, the “material derivative” of \( F \) consists of two
terms: The first term \( \dot{F} \) describes how the cumulative distribution function of the material
element changes with time. This term is also known as the unsteady term. The second
term, \( v \partial_y F \) describes the changes in the cumulative distribution function as the material
element moves from one point to another. This is the \textit{advection} term (convection term for scalar field). For a flow to be \textit{incompressible} (or, isochoric); that is, the velocity field satisfies $\partial_y v = 0$, the sum of these terms should be zero. In other words, if we follow a material element, its cumulative distribution function remains constant. In particular, if the flow is \textit{steady}, then $v \partial_y F = 0$, which shows that $F$ is constant along a streamline. By "steady" we mean the following: a flow in which the velocity of the fluid at a particular fixed point does not change with time, called also \textit{stationary} flow, compare uniform flow.

\subsection*{7.1.2 Lagrangian description of jointly continuous curves}

We present here just the basic case of time–space continuous densities and velocities, while some absolutely continuous of probability densities with respect to the Lebesgue measure at all times are possible. The following statement, which is a first main theorem of this thesis, makes three distinct claims: first, the joint continuity in time and space of the velocity $v$ ensures that the continuity equation (5.5) has at most one solution $\rho \in AC^1(0,T;\mathcal{P}_1^{ac}(\mathbb{R}))$ whose density $\rho = \rho(t,x)$ is jointly continuous in time and space. Second, if such continuous solution exists, then it is also unique within the larger class $AC^1(0,T;\mathcal{P}_1^{ac}(\mathbb{R}))$ (no continuity of densities imposed). Finally, the Lagrangian description of such jointly continuous solution (if it exists) is unique (it is precisely the one provided by Corollary 6.34).

\begin{theorem}
Let $AC_{cont}^1(0,T;\mathcal{P}_1^{ac}(\mathbb{R}))$ be the set of all $\rho \in AC^1(0,T;\mathcal{P}_1^{ac}(\mathbb{R}))$ such that $\rho \in C([0,T] \times \mathbb{R})$. If $v \in C([0,T] \times \mathbb{R})$, then there exists at most one curve $\rho \in AC_{cont}^1(0,T;\mathcal{P}_1^{ac}(\mathbb{R}))$ that originates at a given probability density $\rho_0 \in C(\mathbb{R})$ whose velocity is $v$. Furthermore, if such a curve exists, then it is also the unique $AC^1(0,T;\mathcal{P}_1^{ac}(\mathbb{R}))$ curve starting at $\rho_0$ and whose velocity is $v$. Finally, its only Lagrangian description $X \in W^{1,1}(0,T;L^1(\mathbb{R}))$ starting at a given $X_0$ (such that $X_0 \# \chi = \rho_0$) is given by $X_t = M_t \circ g_0$. Here, $M_t$ are the optimal maps such that $M_t \# \chi = \rho_t$ for all $t \in [0,T]$, and $g_0$ is the almost–everywhere unique $\chi$–preserving map such that $X_0 = M_0 \circ g_0$.

In the statements of the upcoming theorems we used the notation $AC_{cont}^1(0,T;\mathcal{P}_1^{ac}(\mathbb{R}))$ for the space of absolutely continuous curves from $(0,T)$ into $\mathcal{P}^{ac}(\mathbb{R})$ which are continuous in time and space, endowed with the weak and the strong topology, respectively. With
AC^1(0, T; P^{ac} R) we denoted the space of absolutely continuous curves with respect to the Lebesgue measure from (0, T) into P(R), intending that curves in this space are defined for every t ∈ [0, T]. We recall that, up to a redefinition in a negligible set of times, every solution to (5.5) belongs to C_{wa}([0, T]; P(R)) (see Proposition 1.8 and Corollary 1.19).

Proof of Theorem 7.5. Consider the curves ρ ∈ AC^1 cont (0, T; P^{ac} R) and ˜ρ ∈ AC^1 (0, T; P^{ac} R) such that ρ(0, ·) = ρ_0 = ˜ρ(0, ·). Let, as usual, F(t, ·) be the cumulative distribution function of ρ(t, ·). Furthermore, let ˜X ∈ W^{1,1}(0, T; L^1(I)) be a Lagrangian flow map associated with ˜ρ; indeed, we know from Corollary 6.34 that we can, for example, take ˜X ≡ ˜M ◦ g_0, where ˜M is the family of optimal maps such that ˜M_t#χ = ˜ρ_t for all t ∈ [0, T], and g_0 is the χ–preserving map from the polar decomposition of X_0 = M_0 ◦ g_0 = ˜M_0 ◦ g_0 = ˜X(0, ·). Our strategy is to analyze the function

\[ g(t, x) := F(t, ˜X(t, x)), \text{ for all } (t, x) \in [0, T] \times I. \]

We will show that g(·, x) ∈ W^{1,1}(0, T) for almost every x ∈ I and ˙g(·, x) ≡ 0. Since

\[ g(0, x) = F(0, ˜X(0, x)) = ˜F(0, ˜X(0, x)) = g_0(x) \text{ for almost every } x \in I, \]

this will imply

\[ g(t, x) = g_0(x) \text{ for all } t \in [0, T] \text{ and almost every } x \in I. \] (7.7)

Note that the above displayed equalities hold because M_0 = ˜M_0 (since we have ρ(0, ·) = ρ_0 = ˜ρ(0, ·)) and ρ_0 ≪ L^1 (see (6.34)). In fact, (6.34) gives that F(t, M(t, ·)) = 1d almost everywhere in I for all t ∈ [0, T], where M_t are the optimal maps pushing χ forward to ρ(t, ·).

Before we justify (7.7), let us show why that yields the desired thesis. The claims are as follows:

\[ g(t, x) = g_0(x) \text{ for all } t \in [0, T] \text{ and a.e. } x \in I \text{ implies } \rho(t, ·) = ˜\rho(t, ·) \text{ for all } t \in [0, T] \] (7.8)

and

\[ ˜X_t = M_t ◦ g_0 \text{ for all } t \in [0, T], \] (7.9)
That is, $\rho = \tilde{\rho}$ and the Lagrangian flow is necessarily the one consisting of a time–independent rearrangement of the optimal maps pushing $\chi$ forward to $\rho_t$. Indeed, note that (7.7) implies

$$
\tilde{F}_t\#\tilde{\rho}_t = \chi = g_0\#\chi = g_t\#\chi = F_t\#[\tilde{X}_t\#\chi] = F_t\#\tilde{\rho}_t \text{ for all } t \in [0, T].
$$

Since both $F_t$ and $\tilde{F}_t$ are nondecreasing, we infer (by the uniqueness of the optimal map pushing $\tilde{\rho}_t$ forward to $\chi$) $F(t, y) = \tilde{F}(t, y)$ for $\tilde{\rho}_t$–almost every $y \in \mathbb{R}$. Thus, $\partial_y F(t, y) = \partial_y \tilde{F}(t, y)$ for $\mathcal{L}^1$–almost every $y$ in the interior of the support of $\tilde{\rho}_t$; that is, $\rho(t, \cdot) = \tilde{\rho}(t, \cdot)$ Lebesgue almost every in the interior of the support of $\tilde{\rho}_t$. This means that both densities give rise to the same probability (note that the continuity of either density is not necessary here). Finally, since we now know $\tilde{X}_t\#\chi = \rho_t = M_t\#\chi$, we can write $\tilde{X}_t = M_t \circ s_t$ as the polar factorization of $\tilde{X}_t$. Thus,

$$
g_0 = g_t = F_t \circ \tilde{X}_t = F_t \circ M_t \circ s_t = s_t,
$$

which proves claim (7.9).

We now proceed with the proof that $g(\cdot, x)$ is absolutely continuous. Fix $x \in (0, 1)$ for which $t \mapsto \tilde{X}(t, x)$ is in $W^{1,1}(0, T)$, and so we have

$$
\tilde{X}(t, x) = X_0(x) + \int_0^t \dot{\tilde{X}}(s, x) \, ds = X_0(x) + \int_0^t v(s, \tilde{X}(s, x)) \, ds
$$

for $t \in [0, T]$. This implies

$$
|\tilde{X}(t, x)| \leq |X_0(x)| + \|\dot{\tilde{X}}(\cdot, x)\|_{L^1(0, T)} =: C(x) < +\infty \text{ for all } t \in [0, T].
$$

To demonstrate that $t \mapsto g(t, x)$ is absolutely continuous on $[0, T]$ (i.e., $g(\cdot, x) \in W^{1,1}(0, T)$), let us notice first that $g(\cdot, x) \in L^\infty(0, T)$, so it all amounts to proving that there exists $f \in L^1(0, T)$ such that, for all $0 \leq a \leq b \leq T$,

$$
|g(b, x) - g(a, x)| \leq \int_a^b f(t) \, dt.
$$
As expected, we begin by estimating
\[ |g(b, x) - g(a, x)| = |F(b, \tilde{X}(b, x)) - F(a, \tilde{X}(a, x))| \]
\[ \leq |F(b, \tilde{X}(b, x)) - F(b, \tilde{X}(a, x))| + |F(b, \tilde{X}(a, x)) - F(a, \tilde{X}(a, x))| \]
\[ =: E_1 + E_2. \]

We have:
\[ E_1 = \left| \int_{\tilde{X}(a, x)}^{\tilde{X}(b, x)} \rho(b, y) \, dy \right| \leq \left( \max_{[0, T] \times [-C(x), C(x)]} \rho \right) |\tilde{X}(b, x) - \tilde{X}(a, x)|, \]

since \(|\tilde{X}(a, x)|, |\tilde{X}(b, x)| \leq C(x) < +\infty\) and \(\rho \in C([0, T] \times \mathbb{R})\).

Let \(\max \left\{ \max_{[0, T] \times [-R, R]} \rho, \max_{[0, T] \times [-R, R]} |v| \right\} =: \mathcal{M}(R) < +\infty\) for all finite \(R > 0\). Thus,
\[ E_1 \leq \mathcal{M}(C(x)) \int_a^b |\dot{X}(s, x)| \, ds. \tag{7.10} \]

Estimating \(E_2\) is more delicate and requires some regularity of \(t \mapsto F(t, y)\): By using Proposition 7.3 and the observations in Remark 7.4, we have for almost every \(y \in \mathbb{R}\), the map \(F(\cdot, y)\) belongs to \(L^\infty(0, T)\) and has a distributional derivative \(-v(\cdot, y)\rho(\cdot, y) \in L^1(0, T)\). In particular, \(F(\cdot, y) \in W^{1,1}(0, T)\) and
\[ F(b, y) - F(a, y) = - \int_a^b v(t, y)\rho(t, y) \, dt \]
for all \(0 \leq a \leq b \leq T\) and almost every \(y \in \mathbb{R}\).

Since \(\tilde{X}_a \# \chi = \rho_a\) such that \(\rho_a > 0\) implies \(\int_{\mathcal{A}} \rho(t, z) \, dz = \int_{\mathbb{R}} \rho(a, z) \, dz = 1 = \chi(\tilde{X}_a^{-1}(\mathcal{A}))\). Since \(\chi(\tilde{X}_a^{-1}(\mathcal{A})) = 1\) implies \(\tilde{X}_a(x) \in \mathcal{A}\) for almost every \(x \in (0, 1)\), from which we can deduce
\[ F(b, \tilde{X}(a, x)) - F(a, \tilde{X}(a, x)) = - \int_a^b v(t, \tilde{X}(a, x))\rho(t, \tilde{X}(a, x)) \, dt \]
for almost every \(x \in (0, 1)\) and all \(0 \leq a \leq b \leq T\). Thus,
\[ |F(b, \tilde{X}(a, x)) - F(a, \tilde{X}(a, x))| \leq \int_a^b |v(t, \tilde{X}(a, x))|\rho(t, \tilde{X}(a, x)) \, dt \leq \left[ \mathcal{M}(C(x)) \right]^2 (b - a). \tag{7.11} \]

By (7.10) and (7.11), we conclude that for almost every \(x \in (0, 1)\), the function \(g(\cdot, x)\) is absolutely continuous on \([0, T]\).
The next step is to prove that \( \dot{g}(\cdot, x) \equiv 0 \) for almost every \( x \in (0, 1) \). Pick \( t \in (0, T) \) where \( \hat{X}(t, x) \) exists in the pointwise sense, and let \( h \in \mathbb{R} \) such that \(-t/2 \leq h \leq (T - t)/2\).

Set up the difference quotient:

\[
\frac{g(t + h, x) - g(t, x)}{h} = F(t + h, \hat{X}(t + h, x)) - F(t, \check{X}(t, x))
\]

**Case I:** If \( \hat{X}(t + h, x) = \check{X}(t, x) \) for all \( h \) such that \( |h| \leq \delta \) (for some \( \delta > 0 \)), then

\[
\dot{g}(t, x) = \lim_{h \to 0} \frac{F(t + h, \hat{X}(t, x)) - F(t, \check{X}(t, x))}{h} = -\lim_{h \to 0} \frac{1}{h} \int_{t}^{t+h} v(s, \hat{X}(t, x))\rho(s, \check{X}(t, x))ds = -v(t, \hat{X}(t, x))\rho(t, \hat{X}(t, x))
\]

because of the continuity of \( s \mapsto v(s, y)\rho(s, y) \) for all \( y \in \mathbb{R} \). But \( \hat{X}(\cdot, x) \) is constant on \((t - \delta, t + \delta)\), so \( v(t, \hat{X}(t, x)) = \hat{X}(t, x) = 0 \) implies \( \dot{g}(t, x) = 0 \).

**Case II:** There exists a sequence \( \{h_n\}_n \subset [-t/2, (T - t)/2] \) such that

\[
h_n \xrightarrow{n \to +\infty} 0 \quad \text{and} \quad \hat{X}(t + h_n, x) \neq \check{X}(t, x) \quad \text{for all} \quad n \geq 1.
\]

Then,

\[
\frac{g(t + h_n, x) - g(t, x)}{h_n} = \frac{F(t + h_n, \hat{X}(t + h_n, x)) - F(t + h_n, \check{X}(t, x))}{h_n} \frac{\hat{X}(t + h_n, x) - \check{X}(t, x)}{h_n}
\]

\[
= E_1 + E_2.
\]

We have seen that \( E_2 \xrightarrow{n \to +\infty} -v(t, \hat{X}(t, x))\rho(t, \hat{X}(t, x)) \). Thus, it suffices to prove \( E_1 \xrightarrow{n \to +\infty} v(t, \hat{X}(t, x))\rho(t, \hat{X}(t, x)) \). Note that

\[
E_1 = \int_{\hat{X}(t, x)}^{\hat{X}(t + h_n, x)} \rho(t + h_n, y) dy \frac{\hat{X}(t + h_n, x) - \check{X}(t, x)}{h_n}.
\]

Since \( \hat{X}(t, x) \) exists in the pointwise sense (due to our initial choice of \( t \)), we have

\[
\frac{\hat{X}(t + h_n, x) - \hat{X}(t, x)}{h_n} \xrightarrow{n \to +\infty} \hat{X}(t, x) = v(t, \hat{X}(t, x)).
\]
As for \( \int_{X(t,x)}^{\tilde{X}(t+h_n,x)} \rho(t+h_n,y) \, dy \), we use the fact that the restriction of \( \rho \) to \([t - \delta, t + \delta] \times [-C(x), C(x)]\) is uniformly continuous; therefore, \( \tilde{X}(t+h_n,x) \to \tilde{X}(t,x) \) implies
\[
\lim_{n \to +\infty} \int_{X(t,x)}^{\tilde{X}(t+h_n,x)} \rho(t+h_n,y) \, dy = \rho(t, \tilde{X}(t,x)).
\]

This concludes the proof. \( \square \)

Notice that, the last statement of the theorem above basically says that the Lagrangian flow \( X_t \) is a time–independent rearrangement of the optimal maps \( M_t \) because the \( \chi \)–preserving map \( g_0 \) is not depending on the time \( t \) which is, indeed, the identity map.

**Remark 7.6** The assumptions on \( v \) can be weakened, as it can be seen from the proof of Theorem 7.5 above. Indeed, we can only require that \( v \) is locally essentially bounded, \( w := v \rho \) satisfy the conditions from Proposition 7.3, and the map \( \rho v = \rho(t,y)v(t,y) \) is continuous on \([0,T]\) for all \( y \in \mathbb{R} \).

The remark above leads to the following statement:

**Proposition 7.7** Let \( \rho \in AC^1(0,T; \mathcal{P}_p^{ac}(\mathbb{R})) \) for some \( 1 \leq p < +\infty \), and denote by \( v \) the velocity associated to \( \rho \). Assume:

(i) \( \rho \in C([0,T] \times \mathbb{R}) \) and \( \rho > 0 \) everywhere.

(ii) For all \( K \subseteq \mathbb{R} \), there exists \( V_K \in L^1(0,T) \) such that \( |v(\cdot, y)| \leq V_K \) almost everywhere in \((0,T)\) for all \( y \in K \).

If \( p = 1 \), also assume \( v(t, \cdot) \in L^\infty(\rho(t, \cdot)) \) for almost everywhere \( t \in (0,T) \). Then, the Lagrangian flow associated to \( (\rho, v) \) is unique.

**Proof.** Let, as usual, \( g(t,x) = F(t,X(t,x)) \), where \( X \in W^{1,1}(0,T; L^1(I)) \) is a flow map associated with \( v \) and \( F(t, \cdot) \) is the cumulative distribution function of \( \rho(t, \cdot) \) for all \( t \in [0,T] \).
We will show that \( g(\cdot, x) \in W^{1,1}(0,T) \) for almost every \( x \in I \) and \( \dot{g}(\cdot, x) \equiv 0 \), which will imply \( g(t,x) = x \) for almost every \((t,x) \in (0,T) \times I\), that is, \( X \equiv M \) almost everywhere in \((0,T) \times I\). Then, the proof of absolute continuity of \( g(\cdot, x) \) can be carried out exactly as in Theorem 7.5 for the flow \( X \equiv \tilde{X} \). \( \square \)
Observe that, the condition (ii) above is satisfied, for example, by any bounded \( v \). More generally, if \((\exists) \alpha, \beta \in L^1(0, T)\) and \( f \in L^\infty_{\text{loc}}(\mathbb{R})\) such that \( |v(t, y)| \leq \alpha(t)f(y) + \beta(t)\), then \( v \) satisfies (ii). Thus, according to this hypothesis, we have that a map \( t \mapsto v(s, X(t, x)) \) is equiintegrable on \([0, T]\) with respect to \( t \in [0, T]\).

### 7.1.3 The \( L^1 \)–velocity of jointly continuous curves in \( AC^1(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \)

The first assertion of Theorem 7.5 above basically states that, whenever \( v \in C([0, T] \times \mathbb{R}) \), then it gives at most one curve \( \rho \in AC^1_{\text{cont}}(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \); indeed, it does not give a uniqueness for (5.1)–(5.2) for some fixed \( x \in I \), just uniqueness for (5.1)–(5.2) under the constraint \( X_t \# \chi = \rho_t \). The result below shows that for \( v \in C([0, T] \times \mathbb{R}) \) with extra conditions such densities \( \rho \in AC^1_{\text{cont}}(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \) may exist.. Before coming up with an application (Corollary 7.10), we need the following statement, which covers a classical case, as we shall show next.

**Proposition 7.8** Let \( v : [0, T] \times \mathbb{R} \to \mathbb{R} \) satisfy:

(i) \( v \in C([0, T] \times \mathbb{R}) \).

(ii) For all \( t \in [0, T] \), \( v(t, \cdot) \in C^1(\mathbb{R}) \).

(iii) There exists \( \lambda \in L^1(0, T) \) such that \( |\partial_y v(t, y)| \leq \lambda(t) \) for all \( (t, y) \in [0, T] \times \mathbb{R} \).

Then, for any positive probability density \( \rho_0 \in C(\mathbb{R}) \cap \mathcal{P}_1(\mathbb{R}) \), \( v \) is the \( L^1 \)–velocity of a curve \( \rho \in AC^1_{\text{cont}}(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \) originating at \( \rho_0 \).

**Proof.** Let us begin by noticing that \( \rho_0 > 0 \) everywhere implies \( M_0 \) is continuous, strictly increasing on \((0, 1)\), and \( M_0(0+) = -\infty, M_0(1-) = +\infty \). Also, \( M_0 \) is the true inverse of the the cumulative distribution function \( F_0 \) of \( \rho_0 \). This shows that both \( F_0 \in C^1(\mathbb{R}) \) and \( M_0 \in C^1(0, 1) \).

By the classical theory, for each \( x \in I \) the initial value problem (5.1)–(5.2) with \( X_0(x) = M_0(x) \) admits a (unique) solution \( X(t, x) \). Here, we fix \( x \in I \) and \(-x < h < 1 - x, h \neq 0\) and let

\[
Y_h(t, x) := \frac{X(t, x + h) - X(t, x)}{h},
\]
so that it satisfies
\[
\dot{Y}_h(t, x) = \frac{1}{h} [v(t, X(t, x + h)) - v(t, X(t, x))] = f_h(t, x) Y_h(t, x),
\]
where
\[
f_h(t, x) := \int_0^1 \partial_y v(t, (1 - \tau) X(t, x) + \tau X(t, x + h)) d\tau.
\]
Thus,
\[
Y_h(t, x) = \frac{M_0(x + h) - M_0(x)}{h} \exp \left[ \int_0^t f_h(s, x) ds \right],
\]
which gives, in particular,
\[
|X(t, x + h) - X(t, x)| \leq e^{\|\lambda\|_{L^1(0, T)}} |M_0(x + h) - M_0(x)|.
\]
We get from this that \(X(t, \cdot)\) is continuous in \(I\). Next, we have, for all \(\tau \in [0, 1]\),
\[
\lim_{h \to 0} \partial_y v(t, (1 - \tau) X(t, x) + \tau X(t, x + h)) = \partial_y v(t, X(t, x))
\]
by the continuity of \(\partial_y v(t, \cdot)\) and \(X(t, \cdot)\). By (iii), we have that, for \(\mathcal{L}^1\)–almost every \(t \in (0, T)\) and all \(\tau \in [0, 1]\)
\[
|\partial_y v(t, (1 - \tau) X(t, x) + \tau X(t, x + h))|, |\partial_y v(t, X(t, x))| \leq \lambda(t) < +\infty.
\]
Thus, we use the Dominated Convergence Theorem to integrate in \(\tau\) and get
\[
f_h(t, x) \xrightarrow{h \to 0} f(t, x) \quad \text{for almost every } t \in (0, T).
\]
Since \(\lambda \in L^1(0, T)\), we use the Dominated Convergence Theorem again (for the integrals in \(t\) this time) to infer
\[
\lim_{h \to 0} \|f_h(\cdot, x) - f(\cdot, x)\|_{L^1(0, T)} = 0.
\]
Consider now
\[
Y(t, x) := M'_0(x) \exp \left[ \int_0^t \partial_y v(s, X(s, x)) ds \right],
\]
that is, the unique solution of

\[ \dot{Y}(t, x) = \partial_y v(t, X(t, x)) Y(t, x), \quad Y(0, x) = M'_0(x). \] (7.16)

Set \( D_h(t, x) := Y_h(t, x) - Y(t, x) \) to get, according to (7.12) and (7.16),

\[ \dot{D}_h(t, x) = f(t, x) D_h(t, x) + R_h(t, x), \]

where \( f(t, x) := \partial_y v(t, X(t, x)) \) and

\[
R_h(t, x) := Y_h(t, x) \int_0^1 \left[ \partial_y v(t, (1 - \tau)X(t, x) + \tau(t, x + h)) - \partial_y v(t, X(t, x)) \right] d\tau = Y_h(t, x)[f_h(t, x) - f(t, x)].
\]

The solution procedure yields

\[ D_h(t, x) = D_h(0, x) \exp \left[ \int_0^t f(s, x) ds \right] + \int_0^t R_h(s, x) \exp \left[ \int_s^t f(u, x) du \right] ds, \]

which implies \(|D_h(t, x)| \leq e^{\|\lambda\|_{L^1(0,T)}} \left[ |D_h(0, x)| + \int_0^T |R_h(t, x)| dt \right].\) But (7.13) and (7.14) give

\[
\int_0^T |R_h(t, x)| dt = \int_0^T |Y_h(t, x)||f_h(t, x) - f(t, x)| dt \\
\leq e^{\|\lambda\|_{L^1(0,T)}} \left| M_0(x + h) - M_0(x) \right| \|f_h(\cdot, x) - f(\cdot, x)\|_{L^1(0,T)} \xrightarrow{h \to 0} 0
\]

for all \( x \in I \) (since \( M_0 \in C^1(0,1) \)). Since

\[ D_h(0, x) = \frac{M_0(x + h) - M_0(x)}{h} - M'_0(x) \xrightarrow{h \to 0} 0 \text{ for all } x \in I, \]

we deduce \( Y_h(t, x) \xrightarrow{h \to 0} Y(t, x) \) for all \((t, x) \in [0, T] \times I\), which means that, for all \( t \in [0, T] \)

\[ X(t, \cdot) \text{ is differentiable at all } x \in I \quad \text{and} \quad \partial_x X(t, x) = Y(t, x). \] (7.17)

The formula (7.15) for \( Y \) clearly shows (in light of the hypothesis (iii) and the continuity
of $X(t, \cdot)$ that $Y(t, \cdot)$ is continuous in $I$, so we get that $X(t, \cdot) \in C^1(0, 1)$ for all $t \in [0, T]$. Since $M_0$ is strictly increasing in $I$, we also get from (iii) and (7.15) that
\[ e^{-\|\lambda\|_{L^1(0,T)} M'_0(x)} \leq \partial_x X(t, x) \leq e^{\|\lambda\|_{L^1(0,T)} M'_0(x)}, \]
which yields
\[ e^{-\|\lambda\|_{L^1(0,T)} [M_0(y) - M_0(x)]} \leq X(t, y) - X(t, x) \leq e^{\|\lambda\|_{L^1(0,T)} [M_0(y) - M_0(x)]} \] (7.18)
for all $0 < x \leq y < 1$. It immediately follows that $X(t, \cdot)$ is strictly increasing in $I$ and $X(t, 0^+) = -\infty$ and $X(t, 1^-) = +\infty$. If $\rho(t, \cdot) := X(t, \cdot) \# \chi$, we have $\rho(t, X(t, x)) \partial_x X(t, x) = 1$ for all $x \in I$, that is,
\[ \rho(t, y) = \exp \left[ -\int_0^t \partial_y v(s, X(s, F(t, y))) ds \right] \frac{M'_0(F(t, y))}{M'_0(F(t, y))} > 0 \] (7.19)
in light of (7.15) (here, $F(t, \cdot)$ is the true inverse of $X(t, \cdot)$ or, equivalently, the cumulative distribution function of $\rho(t, \cdot)$). Clearly, $\rho(0, \cdot) = \rho_0$.

We claim that $\rho$ defined above belongs to $AC^1_{cont}(0, T; \mathcal{P}_1^{ac}(\mathbb{R}))$, and $v$ is its $L^1$–velocity.

First, we would like to apply Theorem 6.16 to prove $\rho \in AC^1(0, T; \mathcal{P}_1(\mathbb{R}))$ and that $v$ is the required velocity map; according to this theorem, it is enough to prove that $X \in W^{1,1}(0, T; L^1(I))$. Let $m_0 \in \mathbb{R}$ denote the first moment of $\rho_0$, and estimate
\[
\int_{\mathbb{R}} |y| \rho(t, y) dy = \int_I |X(t, x)| dx \leq m_0 + \int_I |X(t, x) - M_0(x)| dx \\
\leq m_0 + \int_0^t \int_I |v(s, X(s, x))| dx ds \\
\leq m_0 + \int_0^t |\lambda(s)| \int_I |X(s, x)| dx ds + \int_0^t |v(s, 0)| ds \\
\leq C_0 + \int_0^t \lambda(s) \int_I |X(s, x)| dx ds,
\]
where $C_0 := m_0 + T \|v(\cdot, 0)\|_{L^{\infty}(0,T)} < +\infty$ (since $t \mapsto v(t, 0)$ is continuous on $[0, T]$). Gronwall’s Lemma now gives a uniform (with respect to $t \in [0, T]$) bound on the first moment of
\[ \rho(t, \cdot) \text{ or, equivalently, on } \|X(t, \cdot)\|_{L^1(I)}. \] Then,
\[
\int_I |\dot{X}(t, x)| \, dx \leq \int_0^T \int_I |v(s, X(s, x))| \, dx \, ds = \int_0^T \int_{\mathbb{R}} |v(s, y)| \rho(s, y) \, dy \, ds
\]
\[
\leq \int_0^T \lambda(s) \int_{\mathbb{R}} |y| \rho(s, y) \, dy \, ds + \int_0^T \int_{\mathbb{R}} |v(s, 0)| \rho(s, y) \, dy \, ds
\]
\[
\leq \|\lambda\|_{L^1(0,T)} \sup_{t \in [0,T]} \|X(t, \cdot)\|_{L^1(I)} + \|v(\cdot, 0)\|_{L^\infty(0,T)} < +\infty.
\]

Thus, \( X \in W^{1,\infty}(0,T; L^1(I)) \subset W^{1,1}(0,T; L^1(I)) \), so \( \rho \in AC^1(0,T; P_1(\mathbb{R})) \) and \( v \) is its \( L^1 \)–velocity map (by Theorem (6.16)).

It only remains to prove that \( \rho \in C([0,T] \times \mathbb{R}) \). The plan is to show first that \( F \) is (jointly) continuous in \([0,T] \times \mathbb{R}\), then use (7.19) to infer that \( \rho \) has the same property. Since \( \rho(t, \cdot) \ll L^1 \) is a probability density, we have that \( y \mapsto F(t,y) \) is \textit{uniformly} continuous in \( \mathbb{R} \), but we would like to show more: namely, that \( F(t, \cdot) \) is uniformly continuous in \( \mathbb{R} \), \textit{uniformly} with respect to \( t \in [0,T] \). For this, denote by \( \omega \) a modulus of continuity for \( F_0 \), i.e. \( \omega : [0, \infty) \to [0, \infty) \) is continuous, increasing, \( \omega(0) = 0 \), and satisfies
\[
\omega(y_2 - y_1) \geq F_0(y_2) - F_0(y_1) \quad \text{for all } y_1 \leq y_2 \in \mathbb{R}.
\]

This is equivalent to
\[
\omega(M_0(x_2) - M_0(x_1)) \geq x_2 - x_1 \quad \text{for all } x_1 \leq x_2 \in I.
\]

Now, let \( \alpha := \exp(\|\lambda\|_{L^1(0,T)}) > 0 \), and use (7.18), the above displayed inequality, and the monotonicity of \( \omega \) to infer
\[
\omega(\alpha[X(t, x_2) - X(t, x_1)]) \geq x_2 - x_1 \quad \text{for all } x_1 \leq x_2 \in I,
\]

That is (upon letting \( \omega_\alpha(r) := \omega(\alpha r) \))
\[
\omega_\alpha(y_2 - y_1) \geq F(t,y_2) - F(t,y_1) \quad \text{for all } y_1 \leq y_2 \in \mathbb{R}.
\]

But \( \omega_\alpha \) is also a modulus of continuity (independent of \( t \in [0,T] \)), so \( F(t, \cdot) \) is uniformly
continuous in $\mathbb{R}$, uniformly with respect to $t \in [0, T]$. The continuity of $[0, T] \ni t \mapsto F(t, y)$ for all $y \in \mathbb{R}$ will also be needed in order to infer that $F$ is continuous in $[0, T] \times \mathbb{R}$. To prove this, assume there exists $y \in \mathbb{R}$ such that $F(\cdot, y)$ is not continuous at some $t \in [0, T]$. Thus, there exists a sequence $\{t_n\}_n \subset [0, T]$ and $\delta > 0$ such that $t_n \to t$ and $|F(t_n, y) - F(t, y)| > \delta$ for all $n \geq 1$; so, it is either $F(t_n, y) - F(t, y) > \delta$ or $F(t, y) - F(t_n, y) > \delta$ for a subsequence (not relabeled) of $\{t_n\}_n$. Assume the former. Since $F(t_n, y) > \delta + F(t, y)$, we deduce $1 > \delta + F(t, y) > 0$, so $\delta + F(t, y)$ is in the domain of $X(t_n, \cdot)$. We know $X(t_n, \cdot)$ is strictly increasing in $I$ for all $n \geq 1$, so $X(t_n, F(t_n, y)) > X(t_n, F(t, y) + \delta)$; that is, $y > X(t_n, F(t_n, y) + \delta)$ for all $n \geq 1$. But $X(\cdot, x)$ is continuous on $[0, T]$ for all $x \in I$ (satisfies (5.1)–(5.2) for all $x \in I$), so we can pass to the limit as $n \to +\infty$ in the last inequality to deduce, after using again that $X(t, \cdot)$ is strictly increasing on $I$,

$$y \geq X(t, F(t, y) + \delta) > X(t, F(t, y)) = y,$$

a contradiction. If, instead, $F(t, y) - F(t_n, y) > \delta$, we rewrite it as $F(t, y) - \delta > F(t_n, y)$, which implies $1 > F(t, y) - \delta > 0$: that is, $F(t, y) - \delta$ is in the domain of $X(t_n, \cdot)$. Since $X(t_n, \cdot)$ is strictly increasing in $I$ for all integers $n \geq 1$, we deduce $X(t_n, F(t, y) - \delta) > X(t_n, F(t, y)) = y$ for all integers $n \geq 1$. As before, in the limit we find

$$y = X(t, F(t, y)) > X(t, F(t, y) - \delta) \geq y,$$

a contradiction. Thus, in light of the uniform continuity of $F(t, \cdot)$ holding uniformly with respect to $t$, we get $F \in C([0, T] \times \mathbb{R})$, as desired.

Thus, if $t_n \to t$ and $y_n \to y$, we can use the continuity of $x \mapsto X(s, x)$ and of $y \mapsto \partial_y v(s, y)$ (the latter, according to (ii)) to deduce

$$\partial_y v(s, X(s, F(t_n, y_n))) \xrightarrow{n \to +\infty} \partial_y v(s, X(s, F(t, y))) \text{ for each } s \in [0, T].$$

We use Dominated Convergence Theorem in light of (iii) to move on to

$$\int_0^t \partial_y v(s, X(s, F(t_n, y_n)))ds \xrightarrow{n \to +\infty} \int_0^t \partial_y v(s, X(s, F(t, y)))ds.$$
But (iii) also gives
\[
\left| \int_t^{t_n} \partial_y v(s, X(s, F(t_n, y_n))) ds \right| \leq \left| \int_t^{t_n} \lambda(s) ds \right| \xrightarrow{n \to +\infty} 0,
\]
which, in light of the previously displayed convergence, implies
\[
\int_0^{t_n} \partial_y v(s, X(s, F(t_n, y_n))) ds \xrightarrow{n \to +\infty} \int_0^t \partial_y v(s, X(s, F(t, y))) ds.
\]
This, along with (7.19) and the fact that \(M'_0\) is continuous on \(I\) (note that \(F(t, y) \in I\) for all \(t \in [0, T]\) and all \(y \in \mathbb{R}\); that is, the values 0 and 1 are achieved only as the asymptotic limits of \(F(t, \cdot)\) at \(-\infty\) and \(+\infty\), respectively), implies the continuity of \(\rho\) in \([0, T] \times \mathbb{R}\). \(\square\)

Notice that, Proposition 7.8(ii) says that \(v\) as a function of \(x\) is \(C^1\) (the space of continuously differentiable functions on \(\mathbb{R}\)) and start at \(\rho_0\). Furthermore, the proposition above gives sufficient conditions on the velocity \(v\) to make \(\rho\) existence within \(AC^1(0, T; \mathcal{P}^{ac}_1(\mathbb{R}))\), and Theorem 7.5 gives the uniqueness, therefore \(v\) is a single curve with respect to \(\rho \in AC^1(0, T; \mathcal{P}^{ac}_1(\mathbb{R}))\).

**Remark 7.9** Note that the assumptions on \(v\) made in the statement of Proposition 7.8 do not include the continuity of \(\partial_y v(t, y)\) in time.

### 7.1.4 Uniqueness of solutions for the continuity equation

A continuous velocity would also give uniqueness for the continuity equation (5.5) within the \(AC^1(0, T; \mathcal{P}^{ac}_1(\mathbb{R}))\) class. In light of Theorem 7.5, Proposition 7.8 gives sufficient conditions on the velocity to render it a single curve velocity with respect to \(\rho \in AC^1(0, T; \mathcal{P}^{ac}_1(\mathbb{R}))\); that is, there is a unique curve \(\rho\) in \(AC^1(0, T; \mathcal{P}^{ac}_1(\mathbb{R}))\) with that velocity. Yet another way to look at it is the result below, which is, indeed, a clear connection between Theorem 7.5 and Proposition 7.8.

**Corollary 7.10** For any \(v\) and \(\rho_0\) satisfying the assumptions in Proposition 7.8, there is a unique solution \(\rho \in AC^1(0, T; \mathcal{P}^{ac}_1(\mathbb{R}))\) of (5.5) with initial \(\rho(0, \cdot) = \rho_0\). Furthermore, this solution lies in \(C([0, T] \times \mathbb{R})\), and is everywhere positive at all times.
Remark 7.11 It is not difficult to construct examples where Theorem 7.5 applies but Proposition 7.8 does not. For this, see example below. In this case, all the conclusions of the above corollary hold even though $v$ is much less regular.

Example 7.12 Take $u : \mathbb{R} \to [1, 2]$ to be continuous and nowhere differentiable (some Weierstrass function); then set, for example,

$$
\rho(t, y) := \eta(t) \frac{u(y)}{y^4 + t + 1},
$$

where $\eta(t)$ normalizes $\rho(t, \cdot)$ to a probability density over $\mathbb{R}$. Note that both $\eta$ and $\dot{\eta}$ are bounded away from zero and infinity on $[0, T]$ for any $0 < T < +\infty$. Also, since $\rho \in C([0, T] \times \mathbb{R})$, in order to prove $\rho \in AC_{cont}^1(0, T; P_{ac}^1(\mathbb{R}))$, we just need to check $\rho \in AC_{cont}^1(0, T; P_{ac}^1(\mathbb{R}))$. Set

$$
g(t, y) := \frac{u(y)}{y^4 + t + 1}, \quad h(t, y) := \frac{u(y)}{(y^4 + t + 1)^2} = \frac{g(t, y)}{y^4 + t + 1},
$$

so that, after some computations, we discover

$$
\partial_t F(t, y) = \frac{H(t, \infty)}{G(t, \infty)} \left[ \frac{G(t, y)}{G(t, \infty)} - \frac{H(t, y)}{H(t, \infty)} \right],
$$

where

$$
G(t, y) := \int_{-\infty}^y g(t, z)dz, \quad H(t, y) := \int_{-\infty}^y h(t, z)dz.
$$

We can easily see that $G(\cdot, \infty)$ and $H(\cdot, \infty)$ are bounded on $[0, T]$ away from zero and infinity. Thus, the integrability of $\partial_t F$ is equivalent to that of

$$
\frac{G(t, y)}{G(t, \infty)} - \frac{H(t, y)}{H(t, \infty)} = F(t, y) - \frac{H(t, y)}{H(t, \infty)}.
$$

But $\tilde{H}(t, \cdot) := H(t, \cdot)/H(t, \infty)$ is the cumulative distribution function of the density

$$
\tilde{h}(t, \cdot) := \frac{h(t, \cdot)}{\int_{\mathbb{R}} h(t, z)dz},
$$

which has uniformly (with respect to $t$) bounded first moment. The same is true about
\[ \rho(t,y) = g(t,y)/\int_{\mathbb{R}} g(t,z)dz \]  Thus,

\[ \int_{\mathbb{R}} |\partial_t F(t,y)| dy \leq \sup_{[0,T]} \left[ \frac{H(t,\infty)}{G(t,\infty)} \right] \| F(t,\cdot) - \tilde{H}(t,\cdot) \|_{L^1(\mathbb{R})} = c W_1(\rho_t, \tilde{h}_t) \leq C < +\infty \]

for all \( t \in [0,T] \), i.e., \( \partial_t F \in L^\infty(0,T;L^1(\mathbb{R})) \subset L^1((0,T) \times \mathbb{R}) \). Thus,

\[ W_1(\rho_s, \rho_t) = \| M_s - M_t \|_{L^1(I)} = \| F(s,\cdot) - F(t,\cdot) \|_{L^1(\mathbb{R})} \leq \int_s^t \| \partial_t F(\tau,\cdot) \|_{L^1(\mathbb{R})} d\tau, \]

(see, e.g., [115] for the equality in the middle), and thus, since \( \partial_t F \in L^\infty(0,T;L^1(\mathbb{R})) \), we deduce \( \rho \in AC^\infty(0,T;P_1(\mathbb{R})) \subset AC^1(0,T;P_{ac}^1(\mathbb{R})) \). Since \( \rho > 0 \) everywhere, we use Proposition 7.3 and (5.6) to reconstruct the velocity

\[ v(t,y) = -\frac{\partial_t F(t,y)}{\rho(t,y)} \in C([0,T] \times \mathbb{R}). \]

Since \( \partial_t F \) is differentiable in \( y \) everywhere in \([0,T] \times \mathbb{R}\) and \( \rho \) is positive and nowhere differentiable in \( y \), we infer \( v \) is not differentiable in \( y \) at each point \((t,y)\) where \( \partial_t F(t,y) \neq 0 \). Fix \( t \in [0,T] \), and assume there are \( y_1 < y_2 \in \mathbb{R} \) such that \( \partial_t F(t,y_1) = \partial_t F(t,y_2) = 0 \), that is, \( F(t,y_i) = \tilde{H}(t,y_i) \) for \( i = 1, 2 \). Thus, there exists \( y_1 < \bar{y} < y_2 \) such that \( \partial_y [F(t,\cdot) - \tilde{H}(t,\cdot)](\bar{y}) = 0 \), that is, \( \rho(t,\bar{y}) = \tilde{h}(t,\bar{y}) \), which is equivalent to \( \bar{y}^4 + t + 1 = G(t,\infty)/H(t,\infty) \). This equation has at most two real solutions, so \( v(t,\cdot) \) is differentiable at at most three points. Thus, \((\rho, v)\) satisfies the assumptions of Theorem 7.5, while \( v \) violates those of Proposition 7.8.

### 7.2 Case of higher integrability

In this section, we present a joint work with A. Tudorascu [21] (see also Section 4 of [70] and [92] for another exposition of these results but in special case). We are able to show some higher integrability for a \( AC \)-path of probability measures on the real line associated to the \( L^1 \)-velocity vector field, with \( q > 1 \), which allow to recover the results of uniqueness of Lagrangian flows presented in the previous section. Moreover, these regularities also allow to prove the uniqueness for solutions to the continuity equation. One of the merits of
this approach is the fact that the whole derivation is purely Lagrangian, in the sense that everything is deduced from the higher integrability and from the definition of “Flow of a Borel map” only (see Definition 5.1), with no mention to the Eulerian side of the problem (compare with the derivation we presented in Subsection 2.3.3). The only drawback lies in the assumption $q > 1$, which can be relaxed a bit (in fact our arguments work under the assumption $\rho_t \ll L^1$ is not necessary, see Subsection 6.3.3), but we are presently not able to cover the case $q = 1$, and this does not allow to reach the general problem of existence for such velocities associated to a given $AC^q(0,T; \mathcal{P}_p(R))$ curve. However, since, once again, the main uniqueness results (and all the previous and remaining results in the thesis) are only concerned with the $\rho_t \ll L^1$ case (for which we get what we need from Theorem 6.33 in the previous chapter), we have decided to investigate thesis other questions in future work. As a consequence of it not really fitting in with the remainder of the thesis, we have thus decided to postpone the study under conditions of the former Theorem 6.35 and the conclusion after it (used to be numbered as Conclusion 6.36). They did not impact the main results in any way, except by providing the convergence in (6.24) (see Lemma 6.25).

7.2.1 The functional derivative of optimal maps in $L^q(0,T; L^p(I))$

We first present a helpful lemma. Observe that the spaces $AC^q(0,T; \mathcal{P}_p(R))$ and $W^{1,q}(0,T; L^p(I))$ equivalent (see Theorem 6.16). In the next lemma, we shall define $\dot{M} \in L^q(0,T; L^p(I))$ as a functional derivative of a map $M \in W^{1,q}(0,T; L^p(I))$. The proof in Lemma 6.25 needs no modification. The philosophy behind this result is that, in some specified sense, $\dot{M}$ can be viewed as almost a classical pointwise time–derivative of $M$. We shall understand its importance in the sequel. Also, since $M \in W^{1,q}(0,T; L^p(I))$, we have that it admits a Borel representative. Equation (7.20) below shows that $\dot{M}$ itself has the property. Throughout the paper we identify both $M$ and $\dot{M}$ with their Borel representatives.

**Lemma 7.13** Let $1 \leq p < +\infty$, $1 < q \leq +\infty$, and let $\rho \in AC^q(0,T; \mathcal{P}_p^{ac}(R))$ with $L^1$–velocity $v$ (assumed to exist if $p = 1$). Denote by $M_t$ the optimal maps pushing $\chi$ forward to
\( \rho_t \) for all \( t \in [0,T] \), and define \( M_t := M_0 \) if \( t < 0 \) and \( M_t := M_T \) if \( t > T \). Then,

\[
\lim_{h \to 0} \int_0^T \left\| \frac{M_{t+h} - M_t}{h} - \dot{M}_t \right\|^q_{L^p(I)} dt = 0. \tag{7.20}
\]

**Proof.** By Theorem 6.33, we have that \( M \in W^{1,q}(0,T;L^p(I)) \), which implies

\[
\left\| \frac{M(t+h, \cdot) - M(t, \cdot)}{h} - \dot{M}(t, \cdot) \right\|_{L^p(I)} \leq \int_t^{t+h} \| \dot{M}(s, \cdot) - \dot{M}(t, \cdot) \|_{L^p(I)} ds,
\]

so

\[
\left\| \frac{M(t+h, \cdot) - M(t, \cdot)}{h} - \dot{M}(t, \cdot) \right\|_{L^p(I)} \xrightarrow{h \to 0} 0 \tag{7.21}
\]

for all \( t \in (0,T) \) that are Lebesgue points for \( [0,T] \ni t \mapsto \dot{M}(t, \cdot) \in L^p(I) \), that is, for almost every \( t \in [0,T] \). Also, if \( f(t) := \| \dot{M}(t, \cdot) \|_{L^p(I)} \) and \( Mf \) its Hardy–Littlewood maximal function, we obtain, for \( t \in (0,T) \) and sufficiently small \( |h| \),

\[
\left\| \frac{M_{t+h} - M_t}{h} - \dot{M}_t \right\|_{L^p(I)} \leq f(t) + Mf(t).
\]

Clearly, the right–hand side is guaranteed to lie in \( L^q(0,T) \) only if \( q > 1 \), since \( f \in L^q(0,T) \) and so is \( Mf \). Thus, if \( 1 \leq p < +\infty \) and \( 1 < q \leq +\infty \), we also get, together by (7.21) and the Lebesgue Dominated Convergence Theorem, that

\[
\int_0^T \left\| \frac{M_{t+h} - M_t}{h} - \dot{M}_t \right\|^q_{L^p(I)} dt \xrightarrow{h \to 0} 0, \tag{7.22}
\]

hence we have shown (7.20), as desired. \( \Box \)

Note that, in order for the above definition of a functional derivative \( \dot{M} \in L^q(0,T;L^p(I)) \) to make sense, we can view \( M \in W^{1,q}(\mathbb{R};L^p(I)) \) as a map by extending \( M_t := M_{0+} \) if \( t < 0 \) and \( M_t := M_{T-} \) if \( t > T \). From here, we can prove just as in [70] that \( M \in W^{1,q}(0,T;L^p(I)) \), then there is a Borel map \( v \) whose flow \( M \) is. Theorem 6.16 then implies \( \rho \in AC^q(0,T;\mathcal{P}_p(\mathbb{R})) \) and \( v \) is its \( L^1 \)–velocity. Therefore, in this case we have that \( M \in W^{1,q}(0,T;L^p(I)) \) implies \( \rho \in AC^q(0,T;\mathcal{P}_p(\mathbb{R})) \). The converse is open.
7.2.2 Uniqueness of the Lagrangian flows in \( W^{1,1}(0, T; L^1(\mathbb{R})) \)

Here, we show that uniqueness of the Lagrangian description of an absolutely continuous curve of probability measures may also be a consequence of some higher integrability enjoyed by the densities in space–time. The reader will note that the full power of Corollary 6.34 was not needed to prove Theorem 7.5 (we only needed existence of a Lagrangian map \( \tilde{X} \) associated to \( \tilde{\rho} \); its explicit nature, provided by Theorem 6.33, was irrelevant to the proof). The theorem below will use both Corollary 6.34 (to provide a Lagrangian description for \( \tilde{\rho} \)) and Theorem 6.33 (applied to \( \rho \)).

We now state unrelated conditions to Theorem 7.5 give similar results. Just as Theorem 7.5, the theorem below makes multiple claims; beside the uniqueness of the Lagrangian description we also have the uniqueness of solutions for (5.5) within a certain class (see Remarks 7.16 and 7.17 below). This is a different kind of result from the previous one in Theorem 7.5 where we do not need to impose any regularity on the velocity \( v \) or the density \( \rho \) (like continuity) is just integrability condition, and the proof is also very different from the proof of the previous one. This kind of disjoint results with respect to the previous one, are not connected.

**Theorem 7.14** Let \( 1 \leq p < +\infty \) and \( 1 < q < +\infty \), and set \( r := q^*(1 + 1/p^*) \), \( s := 1 + p^* \).

If \( \rho \in AC^q(0, T; \mathcal{P}_p(\mathbb{R})) \cap L^r_{\text{loc}}(0, T; L^s(\mathbb{R})) =: \mathcal{U}(p, q) \),

let

\[
S[\rho] := \left\{ \tilde{\rho} \in \mathcal{U}(p, q) : \tilde{\rho}(0, \cdot) = \rho(0, \cdot) \text{ and } (\exists) \lambda \geq 0 \text{ such that } \rho \leq \lambda \tilde{\rho} \text{ or } \tilde{\rho} \leq \lambda \rho \right\} \quad (7.23)
\]

(The inequalities above are to be understood in the \( L^2 \)-almost everywhere sense.) Then, for any \( \rho \in \mathcal{U}(p, q) \), its \( L^1 \)-velocity \( v \) (if \( p = 1 \), we assume \( v \) exists) is the velocity of no other curve in \( S[\rho] \). Furthermore, the only Lagrangian flow map \( X \in W^{1,1}(0, T; L^1(\mathbb{R})) \) associated to some \( \rho \in \mathcal{U}(p, q) \) starting at a given \( X_0 \) (such that \( X_0\#\chi = \rho_0 \)) is given by \( X_t = M_t \circ g_0 \). Here, \( M_t \) are the optimal maps such that \( M_t\#\chi = \rho_t \) for all \( t \in [0, T] \), and \( g_0 \) is the almost–every unique \( \chi \)-preserving map such that \( X_0 = M_0 \circ g_0 \).
Proof. Consider $\rho \in U[p, q]$. Fix some $\tilde{\rho} \in S[\rho]$ such that $\tilde{\rho} \leq \lambda \rho$ for some $\lambda \in \mathbb{R}$ (if such constant does not exist, then there exists $\lambda \in \mathbb{R}$ such that $\rho \leq \lambda \tilde{\rho}$; in this case one only needs to interchange the roles of $\rho$ and $\tilde{\rho}$ in this proof), and assume $\tilde{X} : [0, T] \times I \to \mathbb{R}$ is a flow map associated with $(\tilde{\rho}, v)$ ($\tilde{X}$ exists and lies in $W^{1,q}(0, T; L^p(I))$, according to Corollary 6.34). Let $g : [0, T] \times I \to I$ be defined by

$$g(t, x) := F(t, \tilde{X}(t, x)), \quad \text{where} \quad F(t, y) := \int_{-\infty}^{y} \rho(t, z) \, dz.$$ 

We have $g_{0} = F_{0} \circ \tilde{X}_{0} = F_{0} \circ X_{0}$, which satisfies $g_{0} \# \chi = \chi$. Thus, note that we obtain the desired conclusions by proving that $g$ is time–independent, then using the proved claim (7.8) in the present context. Clearly, $\tilde{X}_{t} \# \chi = \tilde{\rho}_{t}$ implies $F_{t} \# \rho_{t} = g_{t} \# \chi =: \vartheta_{t}$ for all $t \in [0, T]$, which leads to

$$\int_{0}^{1} \zeta(x) \vartheta_{t}(dx) = \int_{0}^{1} \zeta(g_{t}(x)) \, dx = \int_{\mathbb{R}} \zeta(F_{t}(y)) \rho_{t}(y) \, dy$$

$$\leq \lambda \int_{\mathbb{R}} \zeta(F_{t}(y)) \rho_{t}(y) \, dy = \lambda \int_{0}^{1} \zeta(x) \, dx$$

for all nonnegative $\zeta \in C_{c}(I)$. Thus, $\vartheta_{t}$ is a Borel probability such that $\vartheta_{t} \ll \chi$ and (its density) $\vartheta_{t} \in L^{\infty}(I)$ with uniform bound with respect to $t \in [0, T]$. By Theorem 6.33, we know that the equality $\dot{M}(t, y) = v(t, M(t, y))$ between the distributional time–derivative of $M$ (family of optimal maps $M_{t} \# \chi = \rho_{t}$) and $v_{t} \circ M$ holds $L^{2}$–almost everywhere in $(0, T) \times (0, 1)$. Thus, there exists a set of times $\mathcal{T} \subset (0, T)$ of full Lebesgue measure such that for each $t \in \mathcal{T}$ we have $\dot{M}(t, y) = v(t, M(t, y))$ for $\chi$–almost every $y \in (0, 1)$. Let $A_{t}$ denote the set of all such $y$ for a given $t \in \mathcal{T}$. Note that $1 = \vartheta_{t}(A_{t}) = \chi(g_{t}^{-1}(A_{t}))$ (due to $g_{t} \# \chi = \vartheta_{t}$).

For any $x \in g_{t}^{-1}(A_{t})$, we have $g_{t}(x) \in A_{t}$, so we have $\dot{M}(t, g(t, x)) = v(t, M(t, g(t, x)))$. Since $\mathcal{L}^{1}(g_{t}^{-1}(A_{t})) = 1$, we conclude that

For almost every $t \in (0, T)$ we have

$$\dot{M}(t, g(t, x)) = v(t, M(t, g(t, x))) \quad \text{for almost every} \quad x \in I. \quad (7.24)$$
But the support of $\tilde{\rho}_t$ is included in the support of $\rho_t$, which means, because of $\tilde{X}_t\#\chi = \tilde{\rho}_t$,

$$M_t \circ g_t = M_t \circ F_t \circ \tilde{X}_t \equiv \tilde{X}_t \quad \text{Lebesgue almost everywhere in } I. \quad (7.25)$$

Thus, according to (7.24), for Lebesgue almost every $t \in [0, T]$, we have

$$\dot{M}(t, g(t, x)) = v(t, \tilde{X}(t, x)) \quad \text{for Lebesgue almost every } x \in I,$$

which means

$$\dot{X}(t, x) = \dot{M}(t, g(t, x)) \quad \text{for Lebesgue almost every } x \in I. \quad (7.26)$$

Now, fix an arbitrary $\varepsilon > 0$ sufficiently small. By (7.20), we have

$$\lim_{h \to 0} \int_{\varepsilon}^{T-\varepsilon} \left\| \frac{M_{t+h} - M_t}{h} - \dot{M}_t \right\|_{L^p(I)}^q \, dt = 0.$$

Since $g_t\#\chi = \vartheta_t$ and $\vartheta \in L^\infty((0, T) \times I)$, we deduce

$$\lim_{h \to 0} \int_{\varepsilon}^{T-\varepsilon} \left\| \frac{M_{t+h} - M_t}{h} - \dot{M}_t \circ g_t \right\|_{L^p(I)}^q \, dt = 0,$$

which means, in view of (7.25) and (7.26),

$$\lim_{h \to 0} \int_{\varepsilon}^{T-\varepsilon} \left\| \frac{M_{t+h} \circ g_t - \tilde{X}_t}{h} - \dot{\tilde{X}}_t \right\|_{L^p(I)}^q \, dt = 0.$$

But we also have $\tilde{X} \in W^{1,q}(0, T; L^p(I))$, so (7.20) yields

$$\lim_{h \to 0} \int_{\varepsilon}^{T-\varepsilon} \left\| \frac{\tilde{X}_{t+h} - \tilde{X}_t}{h} - \dot{\tilde{X}}_t \right\|_{L^p(I)}^q \, dt = 0.$$

It follows that

$$\lim_{h \to 0} \int_{\varepsilon}^{T-\varepsilon} \left\| \frac{\tilde{X}_{t+h} - M_{t+h} \circ g_t}{h} \right\|_{L^p(I)}^q \, dt = 0. \quad (7.27)$$
Since $\rho_t$ has no atoms, $M_t$ is strictly increasing for all $t \in (0, T)$, and thus

$$g_{t+h}(x) \neq g_t(x) \iff M_{t+h} \circ g_{t+h}(x) \neq M_{t+h} \circ g_t(x)$$

for any $t \in (\varepsilon, T - \varepsilon)$ and any $h$ sufficiently small. Consequently, either

$$\frac{g_{t+h}(x) - g_t(x)}{h} = 0$$

or

$$\frac{g_{t+h}(x) - g_t(x)}{h} = \frac{F_{t+h} \circ \tilde{X}_{t+h}(x) - F_{t+h}(M_{t+h} \circ g_t(x))}{\tilde{X}_{t+h}(x) - M_{t+h} \circ g_t(x)} \frac{\tilde{X}_{t+h} - M_{t+h} \circ g_t(x)}{h},$$

where we have used (7.25) and the fact that $F_t \circ M_t \equiv \text{Id}$ in $I$ for all $t \in [0, T]$ (this is due to the fact that $\rho_t$ being atom–free; i.e., $F_t$ is continuous on $\mathbb{R}$). Thus,

$$\left| \frac{g_{t+h}(x) - g_t(x)}{h} \right| = \left| \int_{M_{t+h} \circ g_t(x)}^{\tilde{X}_{t+h}(x)} \rho_{t+h}(y) \left| \frac{\tilde{X}_{t+h}(x) - M_{t+h} \circ g_t(x)}{h} \right| dy \right|,$$

where we used that $F_{t+h}' = \rho_{t+h}$ and the convention $\int_a^b f = 0$ if $a = b$.

Since $\rho_t \in L^{1+p^*}(\mathbb{R})$ for all $t \in [0, T]$, we deduce it has a maximal function $M\rho_t \in L^{1+p^*}(\mathbb{R})$ such that $\|M\rho_t\|_{L^1(\mathbb{R})} \leq C_p \|\rho_t\|_{L^{p^*}(\mathbb{R})}$ for some finite constant $C_p$ depending only on $p$.

By the definition of the Hardy–Littlewood maximal function, we get

$$\left| \frac{g_{t+h}(x) - g_t(x)}{h} \right| \leq M\rho_{t+h}(\tilde{X}_{t+h}(x)) \left| \frac{\tilde{X}_{t+h}(x) - M_{t+h} \circ g_t(x)}{h} \right|. \quad (7.28)$$

Note that $\tilde{X}_{t+h} \# \chi = \tilde{\rho}_{t+h}$, $s = 1 + p^*$ and $\tilde{\rho}_{t+h} \leq \lambda \rho_{t+h}$ imply (if $p > 1$)

$$\int_I (M\rho_{t+h})^{p^*}(\tilde{X}_{t+h}(x)) \ dx = \int_{\mathbb{R}} (M\rho_{t+h})^{p^*}(y) \tilde{\rho}_{t+h}(y) \ dy \leq \lambda \|\rho_{t+h}\|_{L^s(\mathbb{R})} \left\| (M\rho_{t+h})^{p^*} \right\|_{L^{s^*}(\mathbb{R})}$$

$$= \lambda \|\rho_{t+h}\|_{L^s(\mathbb{R})} \left\| M\rho_{t+h} \right\|_{L^{s^*}(\mathbb{R})} \leq \lambda C_p \|\rho_{t+h}\|_{L^{s^*}(\mathbb{R})} \left. \right|_{L^{s^*}(\mathbb{R})}.$$ 

Thus,

$$\|M\rho_{t+h} \circ \tilde{X}_{t+h}\|_{L^{s^*}(I)} \leq \lambda C_p \|\rho_{t+h}\|_{L^{1+p^*}(\mathbb{R})} \left(1 \right)^{1+1/p^*}, \quad \text{since } s = 1 + p^*.$$ 

This inequality is also obvious if $p = 1$. Consequently, we can apply Hölder’s inequality to
(7.28) to get
\[
\left\| \frac{g_{t+h} - g_t}{h} \right\|_{L^1(I)} \leq \lambda C_p \left\| \rho_{t+h} \right\|_{L^{1+1/p^*}(\mathbb{R})}^{1+1/p^*} \left\| \frac{\bar{X}_{t+h} - M_{t+h} \circ g_t}{h} \right\|_{L^p(I)}.
\]

If \(1 < q < +\infty\), Hölder’s inequality for the time–integral now gives
\[
\int_{\varepsilon}^{T-\varepsilon} \left\| \frac{g_{t+h} - g_t}{h} \right\|_{L^1(I)} dt \leq \lambda C_p \left( \int_{\varepsilon}^{T-\varepsilon} \left\| \rho_{t+h} \right\|_{L^{1+1/p^*}(\mathbb{R})}^{q(1+1/p^*)} dt \right)^{1/q^*} \left( \int_{\varepsilon}^{T-\varepsilon} \left\| \frac{\bar{X}_{t+h} - M_{t+h} \circ g_t}{h} \right\|_{L^p(I)}^q dt \right)^{1/q}.
\] (7.29)

Note that \(\rho \in L^r_{loc}(0, T; L^s(\mathbb{R}))\) implies \(\int_{\varepsilon/2}^{T-\varepsilon/2} \left\| \rho_t \right\|_{L^s(\mathbb{R})}^r dt < +\infty\), so if we introduce \(f(t) := \left\| \rho_t \right\|_{L^s(\mathbb{R})}\), we have \(f \in L^1(\varepsilon/2, T-\varepsilon/2)\). This gives
\[
\lim_{h \to 0} \int_{\varepsilon}^{T-\varepsilon} |f(t+h) - f(t)| dt = 0,
\]
which yields
\[
\lim_{h \to 0} \int_{\varepsilon}^{T-\varepsilon} \left\| \rho_{t+h} \right\|_{L^s(\mathbb{R})}^r dt = \int_{\varepsilon}^{T-\varepsilon} \left\| \rho_t \right\|_{L^s(\mathbb{R})}^r dt < +\infty.
\]

Together with (7.27) and (7.29) and the arbitrariness of \(\varepsilon\), this implies
\[
\lim_{h \to 0} \int_{\varepsilon}^{T-\varepsilon} \left\| \frac{g_{t+h} - g_t}{h} \right\|_{L^1(I)} dt = 0 \quad \text{for all } \varepsilon > 0 \text{ sufficiently small.}
\]

We are trivially led to
\[
\int_0^T \int_I \varphi(t, x) \frac{g_{t+h}(x) - g_t(x)}{h} dx dt \underset{h \to 0}{\longrightarrow} 0 \quad \text{for any } \varphi \in C^1_c((0, T) \times I),
\]
that is,
\[
\int_0^T \int_I \varphi(t, x) g(t, x) dx dt = 0 \quad \text{for all } \varphi \in C^1_c((0, T) \times I),
\]
which implies \(g \in W^{1,\infty}(0, T; L^\infty(I))\) with functional derivative \(\dot{g} \equiv 0\). \(\square\)

In the above theorem we assume that the density \(\rho\) is not just absolutely continuous measure but it is also a function in \(L^1\) because it lies in the time-space \(L^r_{loc}(0, T; L^s(\mathbb{R}))\) (the space of functions that are \(L^r\) locally with respect to time and the space in \(L^s\) where \(r, s\)}
are given in terms of conjugates of \( p, q \), or exactly have these values in the state of Theorem 7.14). So in that case, we have \( \rho \) belongs to the set \( AC^q \) curves with densities (actually measures) intersection with \( L^1 \) space. Furthermore, we can assume \( \rho \in AC^q(0, T; P^{ac}(\mathbb{R})) \) but we do not have to do that, because once we say it is also in \( L^r_{loc}(0, T; L^s(\mathbb{R})) \), this implies \( \rho \) is a function not just a measure because this is space of functions.

Notice that, as in Theorem 7.5, Theorem 7.14 provides a uniqueness result for the continuity equation (5.5) but within a reasonably general class of solutions. In both cases of continuous and higher integrability, the uniqueness results for the Lagrangian flow and the uniqueness results for the continuity equation are obtained concomitantly from our method of proof. Indeed, the result in Theorem 7.14 is not a stronger than the previous result in Theorem 7.5 because it is not related to it and even the proof is different from there where we assume continuity for both of density and velocity but here we assume only integrability of the density.

There is another case, which is not covered by the statement of Theorem 7.14, but follows easily from it.

**Corollary 7.15** Same conclusions as in Theorem 7.14 hold if \( r = s = +\infty \).

**Proof:** Since, for all \( 1 \leq p < +\infty \) and all \( 1 < q < +\infty \), we have \( P_p(\mathbb{R}) \subset P_1(\mathbb{R}) \) (with the inequality \( W_1 \leq W_p \) between the metrics), and since also, obviously, \( L^\infty_{loc}(0, T; L^\infty(\mathbb{R})) \subset L^{r}_{loc}(0, T; L^\infty(\mathbb{R})) \) (where \( r = q^*(1 + 1/p^*) \), as above) we see that Theorem 7.14 applies to the case

\[
\rho \in AC^q(0, T; P_p(\mathbb{R})) \cap L^\infty_{loc}(0, T; L^\infty(\mathbb{R}))
\]

as well, simply as a result of the inclusion \( AC^q(0, T; P_p(\mathbb{R})) \subset AC^q(0, T; P_1(\mathbb{R})) \). \qed

Now, let us say that we have \( \rho \in AC^q(0, T; P_p(\mathbb{R})) \cap L^{r}_{loc}(0, T; L^{s}(\mathbb{R})) =: \mathcal{U}(p, q) \) lies in the class \( S[\rho] \) (the set of all other curves \( \tilde{\rho} \) in \( \mathcal{U}(p, q) \) that start exactly at a given \( \rho_0 \) and also satisfies the comparable). We are looking at all other \( \tilde{\rho} \) starting at the same \( \rho_0 \) such that \( \rho \) is over \( \tilde{\rho} \) that between two constants \( \frac{1}{\lambda} \) and \( \lambda \) for some positive \( \lambda \) (i.e. \( 0 < \frac{1}{\lambda} \leq \frac{\rho}{\tilde{\rho}} \leq \lambda < +\infty \)). For instance; if \( \lambda = 1 \), then we look at these curves \( \tilde{\rho} \) that starting at the same \( \rho_0 \) such that the function \( \rho \) neither bigger twice of \( \tilde{\rho} \) nor less than half of \( \tilde{\rho} \), so something like this; that is,
what is mean they are comparable in values. Therefore, for any \( \rho \) in \( \mathcal{U}(p, q) \), its \( L^1 \)–velocity field is the velocity for no other curve in \( \mathcal{S}[\rho] \). In other words, we say that there is no other curve is comparable that how we said and we can explain this condition given \( \rho \in \mathcal{U}(p, q) \).

This is equivalence to this meaning, well meaning what is says it is comparable in values is same order the ratio is between the two positive constants. This says if we take a curve in \( \mathcal{U} \), then (it is OK with the \( L^1 \)–velocity assume it is exist, if \( p = 1 \)) we will find no other curve it is comparable to it (this is happened because \( \lambda \) is not fixed). We do not need to assume \( \lambda \in \mathbb{R} \), because \( \rho \) and \( \tilde{\rho} \) are probability densities which means that \( \lambda \) can not be negative. If we get a probability density to less or equal to zero, then because \( \tilde{\rho} \) will be nonnegative and \( \lambda \) with negative we will get a probability is not positive, and that is impossible. Because of \( \lambda \) can be just any positive, we can see the nice thing about that is “\( \lambda \) is not fixed” in the definition just says \( \lambda \) exist, so \( \lambda \) could be vary big when we compare \( \lambda \) (could be 1 billion we just say one over billion very close to zero). So, just this is still a vary general result.

In our saying, there is no comparable curve to \( \rho \) that have the same velocity (that is a one statement of Theorem 7.14).

Now, the other statement in the theorem is the only Lagrangian map associated to \( \rho \) in the set \( \mathcal{S}[\rho] \) starting at \( X_0 \) given by a time–independent rearrangement of optimal maps. Note that, we can say only the first statement in Theorem 7.14. However, we can say the both statements, because actually it is an interesting result to get uniqueness not for the flow map \( X \), but also uniqueness for for any two curves in that class sharing with that \( v \). In other words, \( v \) is single curve (SC) but within that class not over all (since it might have more than a one curve, can be a multiple curve (MC)). For each \( \rho \) like this is no other curve in that class will share with that \( v \); that is, if \( \rho \in \mathcal{U}(p, q) \) has a SC velocity \( v \) within the class \( \mathcal{S}[\rho] \), then there is no other \( \rho \) will be a comparable to other in this set that have the same velocity.

Remark 7.16 In the spirit of (7.23) above, here we may define a comparison among all solutions of (5.5) that lie in \( AC^1(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \); in other words, \( \rho, \tilde{\rho} \in AC^1(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \) are said to be comparable if there exists \( \lambda \geq 0 \) such that, for all \( t \in [0, T] \), we have either \( \rho(t, \cdot) \leq \lambda \tilde{\rho}(t, \cdot) \) or \( \tilde{\rho}(t, \cdot) \leq \lambda \rho(t, \cdot) \), \( L^1 \)–almost everywhere in \( \mathbb{R} \). This is reminiscent of
comparison principles used in PDE to establish uniqueness, yet there are major differences. Our \( \lambda \) (in classical studies \( \lambda = 1 \)) accounts for the constraint that \( \rho \) is a probability density at all times and should therefore be at least 1 (\( \lambda = 1 \) implies trivially \( \rho = \tilde{\rho} \)). Also, we do not prove a comparison principle for (5.5) with \( AC^1(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \) solutions. Instead, the above theorem gives us uniqueness of solutions in the same comparison class to the initial–value problem (for any \( \rho_0 \in \mathcal{P}_1(\mathbb{R}) \)) associated with (5.5).

**Remark 7.17** For any \( \rho \in AC^1(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \), we can define an equivalence relation on \( AC^1(0, T; \mathcal{P}_1^{ac}(\mathbb{R})) \) by \( \tilde{\rho} \sim \rho \) if \( \rho(0, \cdot) \equiv \tilde{\rho}(0, \cdot) \) Lebesgue almost everywhere in \( \mathbb{R} \) and if there exists \( 1 \leq \lambda < +\infty \) such that \( \lambda^{-1} \rho \leq \tilde{\rho} \leq \lambda \rho \), \( L^2 \)–almost everywhere in \([0, T] \times \mathbb{R} \). Then, Theorem 7.14 and Corollary 7.15 apply to conclude that, under the given conditions, no two distinct curves in the same equivalence class share the same velocity.

Lastly, it is not clear whether the constraint on \( X_t \) pushing \( \chi \) forward to a specified \( \mu_t \) enjoying certain regularity is necessary. In other words, it may or may not be necessary to impose conditions on \( \mu \) like what we did in Theorems 7.5, 7.14 to get the uniqueness. In fact, we are not sure whether the conditions in the above theorem or in the previous one of the continuity are necessary for the uniqueness because we do not have an example of a velocity shared by two different curves that do not contain atomic measures. The examples of nonuniqueness that we constructed so far, which is presented in Subsection 5.2.3, all satisfy that \( X_t \) have flat spots (or portions); that is, \( \mu_t \) have atoms. Therefore, the atomless assumption maybe the only condition we need to get uniqueness. All we need to do is assuming \( \rho_t \) have no atoms (i.e., we don’t need these big assumptions in Theorems 7.5, 7.14) because we do not have an example where the flow is nonunique and \( \rho \) have no atoms (all of them have atoms). In each time we have a nonunique of the flow in our examples we have flat spots in \( x \) which give us atoms; it means that the optimal maps increasing but \( X \) does not have to be the optimal map (i.e., it does not have to be increasing). So, maybe the nonatomic measure is the only assumption is necessary because we know the uniqueness is true if we have atom but maybe if we impose the conditions that \( \mu_t \) at a time \( t \) is atomless that is enough (we do not know that yet and this is like an open question!). However, the natural question is:
Question 7.18 Is it enough to require that $\mu_t$ have no atoms for the monotone solution to (5.1)–(5.2) to be the only solution?

This kind of question still really needs answering. The nonatomic property of $\mu$ is equivalent to $M_t$ having no flat portions (i.e., no intervals where it is constant), or the cumulative distribution function of $\mu_t$ being continuous.
Chapter 8

Open problems and future work

This thesis deals with uniqueness of order preserving solutions to the Lagrangian flow 
\( t \mapsto X : [0, T] \times I \to \mathbb{R} \) (Definition 5.1)

\[ \dot{X}(t, x) = v(t, X(x, t)), \quad (8.1) \]

where \( X(0, x) = X_0(x) \) is monotone nondecreasing on the interval \( I = [0, 1] \), and \( v \) is a Borel function on \( [0, T] \times I \), as well as on the Eulerian (continuity equation)

\[ \partial_t \mu + \partial_y (\mu v) = 0 \quad \text{on} \quad (0, T) \times \mathbb{R}, \quad (8.2) \]

where \( \mu_0 = X_0 \# \mathcal{L}^1 |_I \). As we mentioned earlier, the conditions stated in the thesis are, essentially, not directly on the vector field \( v \) but on the curve \( [0, T] \ni t \mapsto \mu_t \) of measures in \( \mathcal{P}(\mathbb{R}) \). Here, we would briefly like to remind the reader that it is unclear how much the conditions on \( \mu_t(dx) = \rho(x, t)dx \) on \( \mathbb{R}^1 \) can be relaxed in order for the uniqueness results of both (8.1) and (8.2) to remain true. For example, it is possible, for the corresponding Lagrangian flow to be unique, it is sufficient for \( \rho \) to be absolutely continuous with respect to the Lebesgue measure at all times (or maybe one can require even less: that \( \rho \) has no atoms at all times).

The present study is fundamentally one–dimensional spatially, and a generalization to higher spatial dimensions will require some extra regularity on the velocity and its flow. The article [114] shows that (8.1) will generally not be satisfied by the monotone (rather,
cyclically monotone in higher dimensions) rearrangements regardless of the choice of velocity (as the velocity associated with an absolutely continuous curve of probabilities is not unique, in general, in dimension higher than 1). However, the reference mentioned shows that if the curve consists of absolutely continuous measures, there is one velocity whose Lagrangian flow is the family of optimal maps. Let \( M_t := \nabla P_t \) denote the optimal map (for quadratic cost, see [34]) pushing the Lebesgue measure restricted to the unit cube in \( \mathbb{R}^d \) to the measure \( \rho_t \) on the curve. The cumulative distribution function of \( \rho_t \) is replaced by the gradient of the Legendre transform \( P_t^* \) of the convex function \( P_t \), that is, \( F_t := \nabla P_t^* \). Formally, the fact that \( \nabla P_t^* \circ \nabla P_t = \text{Id} \) in the unit cube leads, in light of \( \dot{M}(t, x) = v(t, M(t, x)) \), to the transport equation

\[
\partial_t F(t, y) + \nabla F(t, y) \cdot v(t, y) = 0. \tag{8.3}
\]

A computation of the type (7.6), with \( F \) as in (8.3) above and \( X \) a flow of \( v \), reveals the same formal result as in the one-dimensional case: that is, that the Lagrangian flow of \( v \) is unique (the one consisting of the optimal maps). It will certainly be interesting to explore under what conditions and to what extent the above findings can be made rigorous, possibly based on recent regularity results on \( P_t^* \) in space [53] and time [9].

In future work, the author would like to address some interesting applications of the theory developed here to spatially monotone solutions of Hamilton–Jacobi equations with non-standard Hamiltonians (sublinear, concave etc) in bounded domains. For example, the equation

\[
\partial_t u + H(\partial_x u) = 0 \tag{8.4}
\]

with \( H(p) = -1/p \) is closely related to Burger’s equation

\[
\partial_t \rho + \partial_x (\rho^2) = 0 \quad \text{in} \quad (0, \infty) \times \mathbb{R}. \tag{8.5}
\]

If \( \rho_0 \) is a probability density, then the latter equation (8.5) has a unique entropy solution \( \rho \) that conserves mass and stays nonnegative at all times (this follows from the closed formula for the viscous approximation, see, e.g., [59]). The generalized inverse of its cumulative distribution function solves the Hamilton–Jacobi equation (8.4) above in some precise sense.
To explain: if $\rho_0$ is, for instance, essentially bounded, then so is $\rho$ at all later times, and we can therefore use Corollary 7.15 to infer existence and uniqueness of the Lagrangian flow associated to $\rho$. Since $v = \rho$ in this case and the Monge–Ampère equation associated with $M(t, \cdot)\#\chi = \rho(t, \cdot)$ is $\rho(t, M(t, x))\partial_x M(t, x) = 1$ (see Remark 4.28), we obtain that $M$ satisfies the Hamilton–Jacobi equation (8.4) above in some precise sense (see Corollary 6.34). Then, properties of $\rho$ as the entropy solution for Burger’s equation (8.5) transfer to $M$ as the solution of the Hamilton–Jacobi equation (8.4) via our theory.

Similarly, we know there is a unique bounded solution $\rho$ for the heat equation

$$
\partial_t \rho - \partial_{xx} \rho = 0 \quad \text{in } (0, +\infty) \times \mathbb{R} \tag{8.6}
$$

if $\rho(0, \cdot) \equiv \rho_0$ is an essentially bounded probability density on $\mathbb{R}$ (see, e.g., [59]). This solution of (8.6) is also smooth and everywhere positive for all $t > 0$, and one can show that $1/\rho$ admits a classical flow (in space) over $[0, 1]$: that is, $M$ solves $\partial_x M(t, x) = 1/\rho(t, M(t, x))$ with the properties $M(t, 0+) = -\infty$, $M(t, 1-) = +\infty$; here, $M(t, \cdot)$ is precisely the optimal map pushing $\chi$ forward to $\rho(t, \cdot)$. Just as discussed above in the case of Burger’s equation (8.5), we infer via Corollary 6.34 that $M$ solves the second–order non–standard Hamilton–Jacobi equation

$$
\partial_t M + \frac{\partial_{xx} M}{(\partial_x M)^2} = 0. \tag{8.7}
$$

We expect that an elegant theory of monotone viscosity solutions for such Hamilton–Jacobi equations (8.4) and (8.7) in bounded domains can be built around the connection between such equations (of the general form $u_t = F(t, x, u_x, u_{xx})$) and the corresponding scalar conservation laws or diffusion equations on the real line. In our work in progress, we are analyzing this connection in much more generality than presented above. We would like to emphasize here that the method of characteristics fails even in these particular cases, and even for “nice”, monotone initial $M_0$. 
Appendix A

A background for technical materials

In this Appendix we collect main objects we are dealing with and state some well known results (in general without proof) which are important in the development of this thesis (for which we refer to [16], [17], [61], [64], [71] and [109] as good general references).

A.1 Ingredients from measure theory

Let \((\mathcal{X}, d)\) be a metric space. We denote by \(\mathcal{B}(\mathcal{X})\) the family of the Borel subsets of \(\mathcal{X}\), by \(\mathcal{M}(\mathcal{X})\) the family of the locally finite Borel measures on \(\mathcal{X}\), by \(\mathcal{M}_+(\mathcal{X})\) the subset of \(\mathcal{M}(\mathcal{X})\) consisting of all nonnegative locally finite Borel measures on \(\mathcal{X}\) and by \(\mathcal{P}(\mathcal{X})\) the family of the Borel probability measures on \(\mathcal{X}\). In a similar fashion we can consider a family of the Borel probability measures \(\mu\) on \(\mathcal{X}\) with finite \(p\)-th moments that we denote by \(\mathcal{P}_p(\mathcal{X})\), i.e. \(\int_{\mathcal{X}} |x|^p \, d\mu(x) < +\infty\). The total variation of a measure \(\mu \in \mathcal{M}(\mathcal{X})\) is denoted by \(|\mu|\). In the case when \(|\mu|(\mathcal{X}) < +\infty\) we say that \(\mu\) has finite mass and we set \(||\mu||_{\mathcal{M}(\mathcal{X})} = |\mu|(\mathcal{X})\). We say that a measure \(\mu\) is concentrated on a Borel set \(B \subset \mathcal{X}\) if \(|\mu|(\mathcal{X}\setminus B) = 0\). We also define the support of \(\mu\) as the smallest (w.r.t. inclusion) closed set defined by \(\text{spt}(\mu) = \{x \in \mathcal{X} : |\mu|(O) > 0\text{ for every open neighborhood }O\text{ of }x\}\) on which \(\mu\) is concentrated. If \(\mu \in \mathcal{M}(\mathcal{X})\) and \(\nu \in \mathcal{M}_+(\mathcal{X})\) we say that \(\mu\) is absolutely continuous with respect to \(\nu\) (and we write \(\mu \ll \nu\)) if \(|\mu|(B) = 0\) for every Borel set \(B \subset \mathcal{X}\) such that \(\nu(B) = 0\). We say that two measures \(\mu, \nu \in \mathcal{M}(\mathcal{X})\) are mutually singular (and write \(\mu \perp \nu\)) if they are concentrated on disjoint Borel sets. If \(\mu \in \mathcal{M}(\mathcal{X})\) and \(B \subset \mathcal{X}\) is a Borel set, the restriction of \(\mu\) to \(B\) is the
measure \( \mu \ll B \in \mathcal{M}(X) \) defined by \((\mu \ll B)(A) = \mu(A \cap B)\) for every Borel set \( A \subset X \).

If \( f : X \to Y \) is a Borel map between two separable metric spaces \( X \) and \( Y \) and \( \mu \in \mathcal{M}(X) \) we denote by \( f_\#\mu \in \mathcal{M}(Y) \) the pushforward of the measure \( \mu \), defined by \( f_\#\mu(B) := \mu(f^{-1}(B)) = \mu(\{x \in X : f(x) \in B\}) \) for every Borel set \( B \subset Y \). Equivalently, this can be characterized by given a summable function \( x \) w.r.t. \( f_\#\mu \) which implies \( x \circ f \) is summable w.r.t. \( \mu \) and then we have the equality \( \int_Y \xi(y) d(f_\#\mu)(y) = \int_X \xi(f(x)) d\mu(x) \) (we often write \( f_\#\mu = \nu \) for a measure on \( Y \)). We denote by \( \mathcal{L}^d \) the Lebesgue measure on \( \mathbb{R}^d \) and with \( \mathcal{L}^k \) the \( k \)-dimensional Hausdorff measure on \( \mathbb{R}^d \). For any subset \( E \) of \( \mathbb{R}^d \) we define the characteristic function \( \chi_E \) (sometimes write \( 1_E \)) as \( \chi_E(x) = 1 \) if \( x \in E \) and 0 otherwise.

A metric space \( X \) is said to be a Polish space if it is a complete and separable. A family \( \mathcal{F} \subset \mathcal{M}_+(X) \) is bounded if there exists a constant \( C \) with \( \|\mu\|_{\mathcal{M}(X)} \leq C \) for each \( \mu \in \mathcal{F} \). We say that a bounded sequence \( \{\mu_k\}_{k=\infty}^\infty \subset \mathcal{M}_+(X) \) is narrowly convergent to \( \mu \in \mathcal{M}_+(X) \) as \( k \to +\infty \) if \( \lim_{k \to +\infty} \int_X f(x) d\mu_k(x) = \int_X f(x) d\mu(x) \) for every \( f \in C_b(X) \) (the space of continuous and bounded real functions defined on \( X \)). We say that a bounded family \( \mathcal{F} \subset \mathcal{M}_+(X) \) is tight if for every \( \epsilon > 0 \) there exists a compact set \( \mathcal{K}_\epsilon \subset X \) such that \( \mu(X \setminus \mathcal{K}_\epsilon) \leq \epsilon \) for every \( \mu \in \mathcal{F} \). The following theorem characterizes the relatively compact subsets with respect to the narrow topology.

**Theorem A.1 (Prokhorov)*** Let \( X \) be a Polish space. A bounded family \( \mathcal{F} \subset \mathcal{M}_+(X) \) is relatively compact with respect to the narrow convergence if and only if it is tight.

Moreover, a necessary and sufficient condition for tightness is the existence of a coercive functional \( \xi : X \to [0, +\infty] \) such that \( \int_X \xi d\mu \leq 1 \) for every \( \mu \in \mathcal{F} \). A family \( \mathcal{F} \) in \( L^1(X, \mu) \) is said to be equiintegrable if for any \( \epsilon > 0 \) there exists a \( \mu \)-measurable set \( B \) with \( \mu(B) < +\infty \) such that \( \int_{X \setminus B} |\xi| d\mu < \epsilon \) for every \( \xi \in \mathcal{F} \), and for any \( \epsilon > 0 \) there exists \( \delta > 0 \) such that a \( \mu \)-measurable set \( A, \mu(A) < \delta \) implies \( \int_A |\xi| d\mu < \epsilon \) for every \( \xi \in \mathcal{F} \). Notice that, for finite measures (e.g. for probability measures), the first assertion is always trivially satisfied by the choice \( B = X \). Therefore, equiintegrability for a family \( \mathcal{F} \in L^1(X, \mu) \), with \( \mu \) finite, reduces to second assertion only. Due to Dunford-Pettis Theorem: if \( \mu \) is a positive \( \sigma \)-finite measure on the measure space \((X, \Sigma)\) and \( \mathcal{F} \) is a bounded family in \( L^1(X, \mu) \), then \( \mathcal{F} \) is weakly (sequentially) compact if and only if \( \mathcal{F} \) is equiintegrable. A rigorous proof of this
statement would require some additional results from measure theory, therefore we only refer to [16] for a proof (precisely, Theorem 1.38 at page 18).

Note that, reading Villani’s books [115], [116] (also see [17]), you could find a different version of Prokhorov’s theorem: a subset \( \mathcal{P} \subset \mathcal{P}(X) \) is relatively (sequentially) compact in \( \mathcal{P}(X) \) if and only if \( \mathcal{P} \) is tight. This version [115] is the one we used in the theses. As consequence of this version, if \( \mu \in \mathcal{P}(X) \), then \( \{\mu\} \) is a tight set in \( \mathcal{P}(X) \). Actually, there is no need to use Theorem A.1 to prove tightness of \( \{\mu\} \). Indeed, in Polish spaces, every probability measure \( \mu \) is regular (see again [28] for a proof), i.e. for any Borel set \( B \), \( \mu(B) = \sup\{\mu(K) : K \text{ compact}, K \subset B\} = \inf\{\mu(O) : O \text{ open}, B \subset O\} \). Therefore, since \( \mu(X) = 1 < +\infty \), we can always find a compact \( K \) whose measure is as close as we want to \( \mu(X) \), and tightness follows.

It is also simple to check that a bounded family \( \mathcal{F} \subset \mathcal{P}(X \times Y) \) is tight if and only if the families of the marginals \( (\pi_X)_\# \mathcal{F} \subset \mathcal{P}(X) \) and \( (\pi_Y)_\# \mathcal{F} \subset \mathcal{P}(Y) \) are tight (see; for instance, the proof of Theorem 3.11). If \( X, Y \) are separable metric spaces, then a measure valued map \( Y \ni y \mapsto \mu_y \in \mathcal{P} \) is said to be a Borel map if \( y \mapsto \mu_y(B) \) is a Borel map for any Borel set \( B \in \mathcal{B}(X) \). Moreover a monotone class argument implies that \( Y \ni y \mapsto \int_X f(x) \, d\mu_y(x) \) is a Borel map for every Borel map \( f : X \to [0, +\infty] \). We also recall a particular case of the disintegration theorem. Let \( \mu \in \mathcal{P}(X \times Y) \), define \( \nu = (\pi_X)_\# \mu \) and assume that \( \nu \in \mathcal{P}(X) \).

Then there exists a Borel family \( \{\mu_x\}_{x \in X} \subset \mathcal{P}(Y) \), which is uniquely determined \( \nu \)-a.e., such that \( \mu = \int_X \mu_x \, d\nu(x) \). Fix a measure \( \mu \in \mathcal{P}(X) \).

We say that a sequence of functions \( \{f_k\}_{k=1}^{\infty} \) defined on \( X \) and with values in \( \mathbb{R}^n \) converges in \( \mu \)-measure to a function \( f \) if \( \lim_{k \to +\infty} \mu \{ \{ x \in X : |f_k(x) - f(x)| > \delta \} \} = 0 \) for every \( \delta > 0 \).

We say that a Borel set \( E \subset \mathbb{R}^d \) is \( \mathcal{H}^n \)-rectifiable if there exist countably many Lipschitz functions \( f_j : \mathbb{R}^n \to \mathbb{R}^d \) such that \( \mathcal{H}^n \left( E \setminus \bigcup_j f_j(\mathbb{R}^n) \right) = 0 \). We finally recall the coarea formula for Lipschitz functions (see for instance Section 2.12 of [16]): for every Lipschitz map \( f : \mathbb{R}^d \to \mathbb{R}^{d-n} \) and every positive Borel function \( \xi : \mathbb{R}^d \to [0, +\infty] \) there holds

\[
\int_{\mathbb{R}^d} \xi \, d\mathcal{L}^d = \int_{\mathbb{R}^{d-n}} \left( \int_{A_k} \xi \, d\mathcal{H}^n \right) \, d\mathcal{L}^{d-n}(k),
\]

where \( J = \sqrt{\det(\nabla f \cdot \nabla f)} \) is the Jacobian of \( f \) and \( A_k = \{ x \in \mathbb{R}^d : f(x) = k \} \), for \( k \in \mathbb{R}^{d-n} \), are the level sets of the function \( f \). If \( \xi \in L^1(\mathbb{R}^d) \) from the latter formula we also deduce

\[
\int_{\mathbb{R}^d} \xi \, d\mathcal{L}^d = \int_{\mathbb{R}^{d-n}} \left( \int_{A_k} (\xi/J) \, d\mathcal{H}^n \right) \, d\mathcal{L}^{d-n}(k).
\]
A.2 Differentiability of Lipschitz continuous functions

A map \( f : \Omega \subset \mathbb{R}^d \to \mathbb{R}^n \) is Lipschitz continuous if there exists a constant \( L \) such that \( |f(x) - f(y)| \leq L|x - y| \) for every \( x, y \in \Omega \). The smallest constant \( L \) such that this inequality holds is called the \textit{Lipschitz constant} of the function \( f \) and is denoted by \( \text{Lip}(f) \). When \( L = \text{Lip}(f) \) we will also say that \( f \) is \( L \)-Lipschitz continuous. A function \( f : \Omega \to \mathbb{R}^n \) is called \textit{locally} Lipschitz continuous if, for each compact \( K \subset \Omega \), there exists a constant \( L_K \) such that \( |f(x) - f(y)| \leq L_K|x - y| \) for every \( x, y \in K \). A function \( f : \mathbb{R}^d \to \mathbb{R}^n \) is differentiable at \( x \in \mathbb{R}^d \) if there exists a linear mapping \( L : \mathbb{R}^d \to \mathbb{R}^n \) such that \( \lim_{y \to x} \frac{|f(y) - f(x) - L(x - y)|}{|x - y|} = 0 \). If such a linear mapping \( L \) exists, it is unique. We call it the derivative of \( f \) at \( x \) and denote it with \( Df(x) \).

\textbf{Theorem A.2 (Rademacher)} A locally Lipschitz continuous function \( f : \mathbb{R}^d \to \mathbb{R}^n \) is differentiable at \( \mathcal{L}^d \)-a.e. \( x \in \mathbb{R}^d \).

For a proof of Rademacher’s theorem, please refer to [61] (precisely, Theorem 2 at page 81). It is simple to check that every real–valued Lipschitz function \( f : \Omega \subset \mathbb{R}^d \to \mathbb{R} \) can be extended to a function \( \tilde{f} : \mathbb{R}^d \to \mathbb{R} \) with \( \text{Lip}(\tilde{f}) = \text{Lip}(f) \) (by extension we mean that \( \tilde{f}|_{\Omega} = f \)). Indeed, it is sufficient to set \( \tilde{f} = f^+ \) or \( \tilde{f} = f^- \), where \( f^\pm(x) = \inf \{ f(y) \pm L|x - y| : y \in \Omega \} \) and \( L = \text{Lip}(f) \). Let us notice that \( f^+ \) and \( f^- \) are respectively the biggest and the smallest extensions. The same result holds also if \( \Omega \) is a metric space. Moreover, arguing componentwise, we deduce that every Lipschitz function \( f : \Omega \subset \mathbb{R}^d \to \mathbb{R}^n \) can be extended to a function \( \tilde{f} : \mathbb{R}^d \to \mathbb{R}^n \) with \( \text{Lip}(\tilde{f}) = \sqrt{k} \text{Lip}(f) \). In fact, for subsets of \( \mathbb{R}^d \), also the following stronger result holds:

\textbf{Theorem A.3 (Kirszbaur)} Let \( f : \Omega \subset \mathbb{R}^d \to \mathbb{R}^n \) be a Lipschitz function. Then there exists an extension \( \tilde{f} : \mathbb{R}^d \to \mathbb{R}^n \) with \( \text{Lip}(\tilde{f}) = \text{Lip}(f) \).

Its proof is considerably more difficult and we mention it for completeness, although it is not strictly necessary in our presentation. We say that a Borel map \( f : \Omega \subset \mathbb{R}^d \to \mathbb{R}^n \) is approximately differentiable at \( x \in \mathbb{R}^d \) if there exists a linear map \( \mathcal{L} : \mathbb{R}^d \to \mathbb{R}^n \) such that the difference quotients \( y \mapsto [f(x + \epsilon y) - f(x)]/\epsilon \) locally converge in measure as as \( \epsilon \downarrow 0 \).
to $\mathcal{L}y$. This is clearly a local property. It is possible to show that, if $f|_K$ is a Lipschitz map for some set $K \subset \mathbb{R}^d$, then $f$ is approximately differentiable at almost every point of $K$. In the following theorem we show a kind of converse of this statement: an approximately differentiable map can be approximated, in the Lusin sense, with Lipschitz maps.

**Theorem A.4** Let $f : \Omega \to \mathbb{R}^n$. Assume that there exists a sequence of Borel sets $B_k \subset \Omega$ such that $\mathcal{L}^d(\Omega \cup \bigcup_k B_k) = 0$ and $f|_{B_k}$ is Lipschitz for any $k$. Then, $f$ is approximately differentiable at $\mathcal{L}^d$–a.e. $x \in \Omega$. Conversely, if $f$ is approximately differentiable at all points of $\Omega' \subset \Omega$, we can write $\Omega'$ as a countable union of sets $B_k$ such that $f|_{B_k}$ is Lipschitz for any $k$ (up to a redefinition of $f$ in an $\mathcal{L}^d$–negligible set).

### A.3 Functions with bounded variation

We say that a function $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ has bounded variation, and we write $f \in BV(\Omega)$, if $f \in L^1(\Omega)$ and the distribution derivative $Df$ is a vector–valued finite measure in $\Omega$. We also introduce the space $BV_{\text{loc}}(\Omega)$ of functions with locally bounded variation as the class of those $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ such that $f \in BV(\tilde{\Omega})$ for every $\tilde{\Omega} \Subset \Omega$. The spaces $BV(\Omega; \mathbb{R}^n)$ and $BV_{\text{loc}}(\Omega; \mathbb{R}^n)$ are defined by requiring $BV$ or $BV_{\text{loc}}$ regularity on each component. The space $BV(\Omega)$ is a Banach space with the norm $\|f\|_{BV(\Omega)} = \|f\|_{L^1(\Omega)} + \|Df\|_{M(\Omega)}$.

When $f \in L^1(\Omega; \mathbb{R}^n)$ we can consider the set of points of approximate discontinuity $S_f$ defined by $\Omega \setminus S_f = \{x \in \Omega : (\exists) z \in \mathbb{R}^n \text{ s.t. } \int_{B_R(x)} |f(y) - z| dy \to 0 \text{ as } R \to 0\}$. Every $x \in \Omega \setminus S_f$ is called a point of approximate continuity. The value $z \in \mathbb{R}^n$ appearing in the definition, when it exists, is unique and is denoted by $\tilde{f}(x)$. It is simple to check that $S_f$ is a Borel set and that the equality $f = \tilde{f}$ holds $\mathcal{L}^d$–a.e. in $\Omega \setminus S_f$. We also introduce the set of points of approximate jump $J_f \subset S_f$: $J_f = \{x \in S_f : (\exists) \nu \in S^{d-1}, (\exists) z^\pm \in \mathbb{R}^n \text{ s.t. } \int_{B_{R+}(x)} |f(y) - z^\pm| dy \to 0 \text{ as } R \to 0\}$, where we have defined the half–balls $B^\pm_{\nu,R}(x) = \{y \in B_R(x) : (y-x) \cdot \nu \geq 0\}$. The triple $(\nu, z^-, z^+)$, when it exists, is unique up to a permutation of $z^-$ and $z^+$ and a change of sign of $\nu$, and is denoted by $((x), f^-(x), f^+(x))$. Notice that $f^-(x) \neq f^+(x)$ when they exist, since $x \in S_f$. It is easy to check that $J_f$ is a Borel set and that $f^\pm$ and $\nu$ can be chosen to be Borel functions in their domain of definition.
For $BV(\Omega; \mathbb{R}^n)$, with $\Omega \subset \mathbb{R}^d$, we can use the Lebesgue decomposition theorem to obtain $Df = D^a f + D^c f$, with $D^a f \ll \mathcal{L}^d$ and $D^s f \perp \mathcal{L}^d$. The measure $D^a f$ is called the absolutely continuous part of the derivative and the measure $D^s f$ is called the singular part of the derivative. The singular part can be decomposed as $D^s f = D^j f + D^c f = D^s f \perp J_f + D^s f \perp (\Omega \setminus J_f)$, where $D^j f$ is the jump part of the derivative and $D^c f$ is the Cantor part of the derivative. The subspace $SBV(\Omega; \mathbb{R}^n) \subset BV(\Omega; \mathbb{R}^n)$ of special function with bounded variation consists of those $f \in BV(\Omega; \mathbb{R}^n)$ with $D^c f = 0$. The following important result relative to the structure of $BV$ functions holds. For the notion of rectifiable set we refer to Appendix A.1.

**Theorem A.5 (Structure of BV functions)** Let $f \in BV(\Omega; \mathbb{R}^n)$. Then, the approximate jump set is countably $\mathcal{H}^{d-1}$-rectifiable, $\mathcal{H}^{d-1}(S_f \setminus J_f) = 0$ and $D^j f = (f^+ - f^-) \otimes \nu \mathcal{H}^{d-1} \setminus J_f$.

The following deep result by Alberti (see [1] and the recent simplified presentation of [52]) says that the rank–one structure of the jump part of the derivative expressed by the above theorem is also shared by the Cantor part.

**Theorem A.6 (Alberti’s rank–one theorem)** Let $f \in BV(\Omega; \mathbb{R}^n)$. Then there exist two Borel functions $\xi : \Omega \to \mathbb{S}^{d-1}$ and $\eta : \Omega \to \mathbb{S}^{k-1}$ such that $D^s f = \xi \otimes \eta |D^s f|$.

In a similar fashion we can define the space of vector fields with bounded deformation in an open set $\Omega \subset \mathbb{R}^d$ by requiring that the symmetric part of the distributional derivative $Ef = \frac{1}{2}(Df + D^t f)$ is a measure: $BD(\Omega; \mathbb{R}^d) = \{ f \in L^1(\Omega; \mathbb{R}^d) : Ef \text{ is a matrix valued finite measure in } \Omega \}$. We refer to [11] and [112] for a presentation of the main properties of this class of vector fields. As in the $BV$ case we can decompose $Ef = E^a f + E^j f + E^c f$ and define the space of special vector fields with bounded deformation as $SBD(\Omega; \mathbb{R}^d) = \{ f \in BD(\Omega; \mathbb{R}^d) : E^c f = 0 \}$.

### A.4 Maximal functions

We recall here the definition of the local maximal function of a locally finite measure and of a locally summable function and we state some well–known properties. If $\mu$ is a (vector–valued) locally finite measure, then, for every $\lambda > 0$, we define the local maximal function of
\( \mu \) as 
\[ \mathcal{M}_\mu(x) = \sup_{0 < R < \lambda} \frac{\mu(B_R(x))}{\mathcal{L}_d(B_R(x))} = \sup_{0 < R < \lambda} \int_{B_R(x)} d\mu(y) \text{ for every } x \in \mathbb{R}^d. \]

When \( \mu = f \mathcal{L}_d \), where \( f \) is a function in \( L^1_{\text{loc}}(\mathbb{R}^d, \mathbb{R}^n) \), we will often use the notation \( \mathcal{M}_\mu f \) for \( \mathcal{M}_\mu \mu \). If \( \lambda > 0 \), then the local maximal function of \( \mu \) is finite for \( \mathcal{L}^d \)-a.e. \( x \in \mathbb{R}^d \) and we have
\[ \int B_\rho(0) \mathcal{M}_\mu f(y) \, dy \leq c_{d, \rho} + c_d \int_{B_\rho+\lambda(0)} |f(y)||\log(2 + |f(y)|)dy. \] For \( p > 1 \) and \( \rho > 0 \) we have the strong estimate \( \int B_\rho(0) (\mathcal{M}_\mu f(y))^p \, dy \leq c_{d, \rho} + c_d \int_{B_\rho+\lambda(0)} |f(y)|^p \, dy \), which is however false for \( p = 1 \). For \( p = 1 \) we only have the weak estimate \( \mathcal{L}^d(\{ y \in B_\rho(0) : \mathcal{M}_\mu f(y) > t \}) \leq \frac{c_d}{t} \int_{B_\rho+\lambda(0)} |f(y)| \, dy \), for every \( t > 0 \). If \( u \in \text{BV}(\mathbb{R}^d) \), then there exists an \( \mathcal{L}^d \)-negligible set \( N \subset \mathbb{R}^d \) such that \( |u(x) - u(y)| \leq c_d|x-y|((\mathcal{M}_\mu Du(x) + \mathcal{M}_\mu Du(y)) \text{ for } x, y \in \mathbb{R}^d \setminus N \) with \( |x-y| \leq \lambda \). We also recall the Chebyshev inequality: for every \( t > 0 \) we have
\[ \mathcal{L}^d(\{|f| > t\}) \leq t^{-1} \int_{\{|f| > t\}} |f(x)| \, dx \leq t^{-1} \mathcal{L}^d(\{|f| > t\})^{1/p'}\|f\|_{L^p(\Omega)}, \text{ where } (1/p) + (1/p') = 1, \text{ and this implies } \mathcal{L}^d(\{|f| > t\})^{1/p} \leq t^{-1}\|f\|_{L^p(\Omega)}. \]

In Chapter 7 we made an extensive use of the following theorem, the so-called Hardy-Littlewood Maximal Theorem, which is related to Chebychev’s inequality. We define the Hardy-Littlewood maximal function \( \mathcal{M}f : \mathbb{R}^d \to [0, +\infty] \) of a locally integrable function \( f : \mathbb{R}^d \to [-\infty, +\infty] \) by the maximum average value that \( f \) can have on open balls of radius \( R \) centered at each point \( x \in \mathbb{R}^d \), i.e. \( \mathcal{M}f(x) = \sup_{R>0}\int_{B_R(x)} |f(y)| \, dy \). If \( f \in L^p(\mathbb{R}^d) \) then the maximal function \( \mathcal{M}f \) is weak \( L^1 \)-bounded and \( \mathcal{M}f \in L^p(\mathbb{R}^d) \). Before stating the theorem more precisely, for simplicity, let \( \{ f > t \} \) denote the set \( \{ x | f(x) > t \} \). Now we have:

**Theorem A.7 (Hardy-Littlewood)** For \( d \geq 1, 1 < p \leq +\infty \), and \( f \in L^p(\mathbb{R}^d) \), there is a constant \( C \) depending on \( p \) and \( d \) such that \( \|\mathcal{M}f\|_{L^p(\mathbb{R}^d)} \leq C\|f\|_{L^p(\mathbb{R}^d)} \) (i.e. strong-type estimate bounded on \( L^p \)). Moreover, if \( f \in L^1(\mathbb{R}^d) \), then there is a constant \( C \) depending only on \( d \) such that for all \( t > 0 \), we have \( \mathcal{L}^d(\{ \mathcal{M}f > t \}) \leq C t^{-1}\|f\|_{L^1(\mathbb{R}^d)}. \)

There are several proofs of this theorem see for example [73] and [81]. In the case \( d = 1 \), for a locally integrable function \( f : \mathbb{R} \to \mathbb{R} \cup \{ \pm \infty \} \) we write \( \mathcal{M}f(x) = \sup_{h>0}\int_{x-h}^{x+h} |f(y)| \, dy. \)
A.5 Vector–valued maps in Banach Spaces

In this section, we introduce some basic tools from Analysis in Banach Spaces. Let $(\mathcal{X}, \| \cdot \|)$ be a real Banach space with dual space $\mathcal{X}^*$. Let $I$ be the bounded, closed interval $[0, T]$ and $0 < T < +\infty$, and consider functions $f : I \to \mathcal{X}$. We will generalize some of the definitions for real–valued functions of a single variable to vector–valued functions [23], [36], [55], [119].

A.5.1 Measurability

If $A \subset I$, let $\chi_A(t)$ denote the characteristic function of $A$. A simple function $f : I \to \mathcal{X}$ is a function of the form $f = \sum_{k=1}^{N} c_k \chi_{A_k}$, where $A_1, \ldots, A_N$ are Lebesgue measurable subsets of $I$ and $c_1, \ldots, c_N \in \mathcal{X}$. A function $f : I \to \mathcal{X}$ is “strongly” measurable if there is a sequence $\{f_k\}_{k=1}^{+\infty}$ of simple functions such that $f_k(t) \to f(t)$ strongly in $\mathcal{X}$ (i.e. in norm) for $t$ a.e. in $I$. Note that, measurability is preserved under natural operations on functions. A function $f : I \to \mathcal{X}$ is said to be weakly measurable if the real–valued function $\langle \xi, f \rangle : I \to \mathbb{R}$ is measurable for every $\xi \in \mathcal{X}^*$. A function $f$ taking values in a Banach space $\mathcal{X}$ is almost separably valued if there is a set $A \subset I$ of measure zero such that $f(I \setminus A)$ is separable, meaning that it contains a countable dense subset. The relationship between weak and strong measurability is given by the following theorem.

**Theorem A.8 (Pettis)** A function $f : I \to \mathcal{X}$ is strongly measurable if and only if it is weakly measurable and almost separably valued.

This theorem reduces the verification of strong measurability to the verification of measurability of real–valued functions. A function $f$ taking values in a Banach space $\mathcal{X}$ is weakly continuous if $\langle \xi, f \rangle : I \to \mathbb{R}$ is continuous for every $\xi \in \mathcal{X}^*$. The space of such weakly continuous functions is denoted by $C_{ws}(I; \mathcal{X})$. Since a continuous function is measurable, every almost separably valued, weakly continuous function is strongly measurable (see Example 6.12).
A.5.2 Bochner and weak-*integral

Let \( f = \sum_{k=1}^{N} c_k \chi_{A_k} \) be the simple function as defined in above. The integral of \( f \) is defined by \( \int_{0}^{T} f \, dt = \sum_{k=1}^{N} c_k |A_k| \in \mathcal{X} \), where \( |A_k| \) denotes the Lebesgue measure of \( A_k \). A strongly measurable function \( f : I \rightarrow \mathcal{X} \) is Bochner integrable (or, integrable) if there is a sequence of simple functions such that \( f_k(t) \rightarrow f(t) \) pointwise a.e. in \( I \) and \( \lim_{k \rightarrow +\infty} \int_{0}^{T} \|f_k - f\| \, dt = 0 \). The integral of \( f \) is defined by \( \lim_{k \rightarrow +\infty} \int_{0}^{T} f_k \, dt = \int_{0}^{T} f \, dt \), where the limit exists strongly in \( \mathcal{X} \). If \( f : I \rightarrow \mathcal{X} \) is integrable and \( \xi \in \mathcal{X}^* \), then \( \langle \xi, f \rangle : I \rightarrow \mathbb{R} \) is integrable and \( \int_{0}^{T} \langle \xi, f \rangle \, dt = \langle \xi, \int_{0}^{T} f \, dt \rangle \). In fact, \( \xi \) is said to be weakly-* \( \mu \)-integrable (or Gelfand integrable) and, by a simple application of closed graph theorem we have that the maps \( \xi_f(\omega) := \langle \xi(\omega), f \rangle \), are \( \mu \)-measurable whenever a measure space \((\Omega, \Sigma, \mu)\) is defined on a separable Banach space \( \mathcal{X} \). The following result, due to Bochner, characterizes integrable functions as ones with integrable norm.

**Theorem A.9** A function \( f : I \rightarrow \mathcal{X} \) is Bochner integrable if and only if it is strongly measurable and \( \int_{0}^{T} \|f\| \, dt < +\infty \).

Thus, in order to verify that a measurable function \( f \) is Bochner integrable one only has to check that the real valued function \( \|f\| : I \rightarrow \mathbb{R} \), which is necessarily measurable, is integrable (see again Example 6.12). Among all the basic properties of Bochner integral we recall the useful result: if \( f \) is Bochner integrable on \( I = [0, T] \) with respect to Lebesgue measure, then for almost every \( s \in I \) one has \( \lim_{h \rightarrow 0} \int_{s}^{s+h} f(t) - f(s) \, dt = 0 \), and consequently for almost every \( s \in I \)\( \lim_{h \rightarrow 0} \int_{s}^{s+h} f(t) \, dt = f(s) \). The dominated convergence theorem holds for Bochner integrals: suppose that \( f_k : I \rightarrow \mathcal{X} \) is Bochner integrable for each \( k \in \mathbb{N} \), \( f_k \rightarrow f(t) \) as \( k \rightarrow +\infty \) strongly in \( \mathcal{X} \) for a.e. \( t \in I \), and there is an integrable function \( f_k : I \rightarrow \mathbb{R} \) such that \( \|f_k(t)\| \leq g(t) \) for a.e. \( t \in I \) and every each \( k \in \mathbb{N} \). Then, \( f : I \rightarrow \mathcal{X} \) is Bochner integrable and \( \int_{0}^{T} f_k \, dt \rightarrow \int_{0}^{T} f \, dt \) and \( \int_{0}^{T} \|f_k - f\| \, dt \rightarrow 0 \) as \( k \rightarrow +\infty \). The proof is the same as for the scalar–valued case, and we omit it.

For \( 1 \leq p < +\infty \) the space \( L^p(0, T; \mathcal{X}) \) consists of all strongly measurable functions \( f : I \rightarrow \mathcal{X} \) such that \( \int_{0}^{T} \|f\|^p \, dt < +\infty \) equipped with the norm \( \|f\|_{L^p(0, T; \mathcal{X})} = \left( \int_{0}^{T} \|f\|^p \, dt \right)^{1/p} \). The space \( L^\infty(0, T; \mathcal{X}) \) consists of all strongly measurable functions \( f : I \rightarrow \mathcal{X} \) such that
\[\|f\|_{L^\infty(0,T;X)} = \sup_{t \in I} \|f\| < +\infty,\] where sup denotes the essential supremum. We regard functions that are equal pointwise a.e. as equivalent, and identify a function that is equivalent to a continuous function with its continuous representative. If \(X\) is a Banach space and \(1 \leq p \leq +\infty\), then \(L^p(0,T;X)\) is a Banach space. If \(X\) is a Banach space and \(1 \leq p < +\infty\), then the collection of the form \(f = \sum_{i=1}^k c_i \xi_i(t)\), where \(\xi \in C^\infty_c(I)\) and \(c_i \in X\) is dense in \(L^p(0,T;X)\) (here \(\xi\) considered as the mollification of simple functions). The characterization of the dual space of a vector–valued \(L^p\)–space is analogous to the scalar–valued case, after we take account of duality in the range space \(X\). If \(1 \leq p < +\infty\) and \(X\) is a reflexive Banach space with dual space \(X^*\), then the dual of \(L^p(0,T;X)\) is isomorphic to \(L^{p'}(0,T;X^*)\) where \(1/p + 1/p' = 1\). The action of \(f \in L^p(0,T;X)\) on \(\xi \in L^{p'}(0,T;X')\) is given by \(\langle \langle \xi, f \rangle \rangle = \int_0^T \langle \xi(t), f(t) \rangle \, dt\). The proof of this statement is more complicated than in the scalar case and some condition on \(X\) is required. Reflexivity is sufficient (as is the condition that \(X^*\) is separable). We say that \(\xi \in L^{p,*}_w(\mu;X)\) if \(\xi\) is weakly-* \(\mu\)-integrable and \(\int_0^T \|\xi\|^p \, dt < +\infty\).

### A.5.3 Differentiability

The definition of continuity and pointwise differentiability of vector–valued functions are the same as in the scalar case. A function \(f : I \to X\) is strongly continuous at \(t \in I\) if \(f(s) \to f(t)\) strongly in \(X\) as \(s \to t\), and \(f\) is strongly continuous in \(I\) if it is strongly continuous at every point of \(I\). A function \(f\) is strongly differentiable at \(t \in I\), with strong pointwise derivative \(\dot{f}(t)\), if \(\dot{f}(t) = \lim_{h \to 0} [f(t + h) - f(t)]/h\), where the limit exists strongly in \(X\), and \(f\) is continuously differentiable in \(I\) if its pointwise derivative exists for every \(t \in I\) and \(\dot{f} : I \to X\) is a strongly continuously function.

Let \(L^1_{\text{loc}}(0,T;X)\) denote the space of measurable functions \(f : I \to X\) that are integrable on every compactly supported interval \((a,b) \subseteq I\). Also, as usual, let \(C^\infty_c(I)\) denote the space of smooth, real–valued functions \(\xi : I \to \mathbb{R}\) with compact support, \(\text{spt}(\xi) \subseteq I\). A function \(f \in L^1_{\text{loc}}(0,T;X)\) is weakly differentiable with weak derivative \(\dot{f} = g \in L^1_{\text{loc}}(0,T;X)\) if \(\int_0^T \xi f \, dt = -\int_0^T \xi g \, dt\) for every \(\xi \in C^\infty_c(I)\). These integrals are understood as Bochner integrals. The same result of the Lebesgue differentiation theorem in the case of scalar–valued integrals is true for the case of vector–valued integrals.
Theorem A.10 Suppose that $X$ is a Banach space and $f \in L^1(0,T;X)$, then the limit
$$\lim_{h \to 0} \int_{t}^{t+h} f(s) \, ds$$
exists and is equal to $f(t)$ for $t$ pointwise a.e. in $I$.

As consequence, if $f : I \to X$ is locally integrable and $\int_{0}^{T} \xi \, dt = 0$ for every $\xi \in C_c^\infty(I)$, then, $f = 0$ pointwise a.e. on $I$. This corollary corresponds to the statement that a regular distribution determines the values of its associated locally integrable function pointwise almost everywhere. We also have a vector–valued analog of a scalar–valued that the only functions with zero weak derivative are the constant functions. The proof is similar: if $f : I \to X$ is weakly differentiable and $\dot{f} = 0$, then $f$ is equivalent to a constant function. It also follows that a function is weakly differentiable if and only if it is the integral of an integrable function: if $X$ is a Banach space and $f \in L^1(0,T;X)$, then $f$ is weakly differentiable with integrable derivative $\dot{f} = g \in L^1(0,T;X)$ if and only if $f(t) = c_0 + \int_{0}^{T} g(s) \, ds$ pointwise a.e. in $I$. In that case, $f$ is differentiable pointwise a.e. and its pointwise derivative coincides with its weak derivative. We can also characterize the weak derivative of a vector–valued function in terms of weak derivatives of the real–valued functions obtained by duality: if $X$ is a Banach space with dual $X^*$ and $f, g \in L^1(0,T;X)$, then $f$ is weakly differentiable with weak derivative $\dot{f} = g$ if and only if, for every $\phi \in X^*$, $\frac{d}{dt} \langle \xi, f \rangle = \langle \xi, g \rangle$ as a real–valued weak derivative in $I$.

We define Sobolev spaces of vector–valued functions in the same way as for scalar–valued functions, and they have similar properties. Suppose that $X$ is a Banach space, $k \in \mathbb{N}$, and $1 \leq p \leq +\infty$. The Banach space $W^{k,p}(0,T;X)$ consists of all (equivalence classes of) measurable functions $u : I \to X$ whose weak derivatives of order $0 \leq j \leq k$ belong to $L^p(0,T;X)$. If $1 \leq p < +\infty$, then the $W^{k,p}$–norm is defined by $\|u\|_{W^{k,p}(0,T;X)} = \left( \int_{0}^{T} \sum_{j=1}^{k} \|\partial_t^j u\|_X^p \, dt \right)^{1/p}$. If $p = +\infty$, then $\|u\|_{W^{k,\infty}(0,T;X)} = \sup_{1 \leq j \leq k} \|\partial_t^j u\|_X$. If $p = 2$, and $X = H$ is a Hilbert space, then $W^{k,2}(0,T;H) = H^k(0,T;H)$ is the Hilbert space with inner product $(u,v)_{H^k(0,T;H)} = \int_{0}^{T} (u(t),v(t))_H \, dt$. The Sobolev embedding theorem for scalar–valued functions of a single variable carries over to the vector–valued case: if $1 \leq p \leq +\infty$ and $u \in W^{k,p}(0,T;X)$, then $u \in C(I;X)$. Moreover, there exists a constant $C = C(p,T)$ such that $\|u\|_{L^\infty(0,T;X)} \leq C \|u\|_{W^{1,p}(0,T;X)}$. 

Mohamed H. Amsaad Appendix A 258
A.5.4 Absolutely continuous curves in metric spaces

Although we use this discussion everywhere throughout this thesis, it is interesting to consider the relationship between weak differentiability and absolute continuity in the vector-valued case. The definition of absolute continuity of vector-valued functions is a natural generalization of the real-valued definition. During the thesis, we study explicit characterizations of the (potential) absolute continuity property of the inverse distribution functions. This property will always be understood in the local sense. Namely, a function \( f \) defined on an interval \( I = [0, T] \), finite or not, will be called absolutely continuous if for every \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that, for any collection \( \{[t_0, t_1], [t_2, t_3], \ldots, [t_{N-1}, t_N]\} \) of non-overlapping subintervals of \( I \), \( \sum_{k=1}^{N}|t_k - t_{k-1}| < \delta \) implies \( \sum_{k=1}^{N}\|f(t_k) - f(t_{k-1})\| < \varepsilon \).

Obviously, every absolutely continuous map is obviously uniformly continuous, and therefore it can be uniquely extended to the closure of \( I \). Similarly, \( f : I \rightarrow \mathcal{X} \) is Lipschitz continuous on \( I \) if there exists a constant \( L \geq 0 \) such that \( \|f(t) - f(s)\| \leq L|t - s| \) for every \( s, t \in I \). It follows immediately that a Lipschitz continuous function is absolutely continuous (with \( \delta = \varepsilon/L \)).

Equivalently, for some locally integrable function \( g \) on \( I \), for all \( 0 \leq t_0 < t_1 \leq T \), \( f(t_1) - f(t_0) = \int_{t_0}^{t_1} g(t) \, dt \). The function \( g \) is uniquely determined up to a set of measure zero, is denoted by \( f' \), and is called the Radon-Nikodym derivative of \( f \). As a possible variant, one may put \( f'(t) = \limsup_{\varepsilon \to 0} (f(t + \varepsilon) - f(t))/\varepsilon \) or use \( \lim \inf \) instead. For example, any locally Lipschitz function is absolutely continuous. Such functions have Radon-Nikodym derivatives that are bounded on proper subintervals of \( I \).

If an absolutely continuous function \( f \) is nondecreasing, the Radon-Nikodym derivative of \( f \) can always be chosen to be nonnegative and a.e. finite. Any such \( f \) generates a nonnegative Borel measure \( \mu(A) = \int_A f'(t) \, dt \), for every a Borel set \( A \subset I \), which is absolutely continuous with respect to the Lebesgue measure on \( I \) in the usual sense of Measure Theory. We do not require that \( \mu \) be finite, but \( \mu \) should be finite on compact subintervals of \( I \). Note that any nonnegative absolutely continuous measure \( \mu \), which is finite on compact subintervals of \( I \), is generated by some nondecreasing absolutely continuous function \( f \).

The space \( AC^p(I; \mathcal{X}), 1 \leq p \leq +\infty \) consists of curves \( f : I \to \mathcal{X} \) such that, for some
function $\beta \in L^p(I)$,

$$d(f(s), f(t)) \leq \int_s^t \beta(\tau) \, d\tau \quad \text{for every } s \leq t \text{ in } I. \quad (A.1)$$

Since $X$ is complete, $\lim_{t \to 0^+} f(t)$ exists and will be denoted $f(0)$. Similarly, $f(T)$ is well-defined. A curve $f \in AC^1(I; X)$ is called absolutely continuous in $X$, and a curve $f \in AC^p(I; X)$, for $p > 1$, is called absolutely continuous with finite $p$-energy. We can also define the space of locally absolutely continuous functions on $\mathbb{R}$ by $AC^p_{loc}(\mathbb{R}; X) = \{ f : \mathbb{R} \to X : f \in AC^p(I; X) \}$. The elements of $AC^p(I; X)$ satisfy the nice property of a.e. metric differentiability. Following the simple argument in [3], [17] that, for any metric space $(X, d)$ and any absolutely continuous map $f : I \subset \mathbb{R} \to X$, the metric derivative of $f$ exists for $L^1$-a.e. times in $I$, and it is the smallest function in $L^p(I)$ satisfying the condition (A.1) above. This is a kind of metric version of Rademacher’s theorem, see also [18] and the references therein for the extension to maps defined on subsets of $\mathbb{R}^d$. Precisely we have the following Theorem (see [17] for the proof).

**Theorem A.11 (Metric derivative)** Let $1 \leq p \leq +\infty$. Then for every curve $f \in AC^p(I; X)$ the limit

$$|f'| (t) := \lim_{h \to 0} \frac{d(f(t + h), f(t))}{|h|} \quad (A.2)$$

exists for $L^1$-a.e. $t \in I$. The function $t \mapsto |f'| (t)$ belongs to $L^p(I)$ and $d(f(s), f(t)) \leq \int_s^t |f'| (\tau) \, d\tau$ for any interval $(s, t) \subset I$. Moreover, for every function $\beta$ satisfying (A.1), $|f'| (t) \leq \beta(t)$ for $L^1$-a.e. $t \in I$.

### A.5.5 The Radon-Nikodým property

Let $\Sigma$ be a $\sigma$-algebra of the set $\Omega$. A vector measure with values in a Banach space $X$ is a countably additive application $F : \Sigma \to X$. A vector measure $F$ is said to be absolutely continuous with respect to a nonnegative finite real valued measure $\mu$ on $\Sigma$ if $\lim_{\mu(B) \to 0} F(B) = 0$. The variation of a vector measure is defined by $|F|(B) = \sup \{ \sum_{j=1}^k \| E_j \| : \{ E_j \} \subset \Sigma \text{ finite partition of } B \}$. The vector measure is said to be of bounded variation if $|F|(\Omega) < +\infty$. A Banach space $X$ has the **Radon-Nikodým property** with respect to a measure space
if for every vector measure $F : \Sigma \to \mathcal{X}$ of bounded variation, absolutely continuous with respect to $\mu$ there exists $\xi \in L^1(\mu; \mathcal{X})$ such that $F(B) = \int_B \xi \, d\mu$ for every $B \in \Sigma$. A Banach space $\mathcal{X}$ has the Radon-Nikodým property if $\mathcal{X}$ has the Radon-Nikodým property with respect to every finite measure space.

We recall that if $\mathcal{X}$ has the Radon-Nikodým property with respect to a Lebesgue measure on $[0, 1]$ then $\mathcal{X}$ has the Radon-Nikodým property. Other equivalent formulations of the Radon-Nikodým property and many examples of spaces that have or don’t have Radon-Nikodým property can be founded in [55]. Here we recall only that reflexive Banach spaces, separable dual spaces, $l^1$, $L^p(\mu; \mathcal{X})$ for $p \in (1, +\infty)$ when $\mathcal{X}$ has the Radon-Nikodým property, have the Radon-Nikodým property. The spaces $L^1(\mu)$, with $\mu$ not purely atomic, $l^\infty$, $L^\infty([0, 1])$, $C(\Omega)$ where $\Omega$ is a bounded open subset of $\mathbb{R}^d$, do not have the Radon-Nikodým property.
Appendix B

Geometrical properties for transports

In this Appendix we analyze properties of $c$-concave functions in comparison with convex functions, which are better known. We start from recalling some notions and few results in convex analysis that we need in the following ingredients. These notions have been written mainly to fill the gaps I left in the previous section of the results contained in Chapter 2 of [115]. Many preliminaries are piled up in this section before a correct statement of the main results of [115] could be given. I tried to select the results in [115] to help readers know about the preliminaries, and I put here most of the tools needed for these results. Anyway, not all the preliminaries are needed to read (or to prove) each result.

Notice that also not all the material here comes from [115]. I also used [116] and, here and there, added different definitions or proofs to keep the consistency of the notations. Anyway, no proof, example or remark is original: I predated many well–written books, papers and notes to collect the results I needed. Take a look at the Bibliography, if you want to know where my arguments come from.

B.1 Ingredients from convex analysis

Let $X$ be a vector space. A function $\xi : X \rightarrow \mathbb{R} \cup \{+\infty\}$, not identically $+\infty$, is said to be a convex function if, for every $x, y \in X$ and $t \in [0, 1]$, we have $\xi(tx + (1 - t)y) \leq t\xi(x) + (1 - t)\xi(y)$. Moreover, $\xi$ is said to be strictly convex if the previous inequality implies either $x = y$ or $t \in \{0, 1\}$. If $t = 1/2$, this definition is equivalent to $\xi((x + y)/2) \leq (1/2)[\xi(x) + \xi(y)]$. 
Finally, $\xi$ is said to be proper if there exists $x \in \mathcal{X}$ such that $\xi(x) < +\infty$. Equivalently, setting $\text{Dom}(\xi) := \{x \in \mathcal{X} : \xi(x) < +\infty\}$, $\xi$ is proper if $\text{Dom}(\xi) \neq \emptyset$. We write $\mathcal{D}(\xi) = \text{Dom}(\xi)$ and $\mathcal{D}^0(\xi) = \text{Int}(\mathcal{D}(\xi))$. Whenever $\xi$ is convex, $\mathcal{D}(\xi)$ is a convex set. In particular, when $\mathcal{X} = \mathbb{R}^d$, since any convex set $C$ has $\partial C$ with Hausdorff dimension $d - 1$ (this should be a consequence of a result contained in [63], but we haven’t tried to obtain it in detail), $\partial \mathcal{D}(\xi)$ is a set with finite $\mathcal{H}^{n-1}$ measure, where $\mathcal{D}(\xi) = \{x \in \mathbb{R}^d : \xi(x) \in \mathbb{R}\}$. We also mention the next equivalent definitions of a convex function. You can prove them as an exercise or look for them in [86], [88], [89], [97], [99], [100].

A function $\xi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if the epigraph (or supergraph) of $\xi$, defined as the set of points lying on or above its graph $\text{Epi}(\xi) := \{(x, \lambda) \in \mathcal{X} \times \mathbb{R} : \xi(x) \leq \lambda\}$, is a convex set. If $\mathcal{X} = \mathbb{R}$, both conditions are also equivalent to, for every $x_o \in \mathbb{R}$, the map $x \mapsto (\xi(x) - \xi(x_o))/(x - x_o)$, is nondecreasing for every $x \neq x_o$. If $\mathcal{X}$ is a normed space and $\xi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a convex map, then $\xi$ is proper, $\mathcal{D}^0(\xi) \neq \emptyset$ and $\xi$ is continuous in $\mathcal{D}^0(\xi)$ if and only if there exist $\mathcal{O} \subset \mathcal{X}$, open and nonempty, and $\alpha \in \mathbb{R}$ such that $\xi(x) \leq \alpha$ for every $x \in \mathcal{O}$. If $x_o \in \mathcal{X}$, $r_o > 0$, $m, M \in \mathbb{R}$ are such that $m \leq \xi(x) \leq M$ for every $x \in B(x_o, r_o)$, then $\xi$ is Lipschitz continuous on $B(x_o, r_o)$, for all $0 < r < r_o$, with Lipschitz constant $\text{Lip}(\xi) := (M - m)/(r_o - r)$. If $\mathcal{X}$ is a normed space and $\xi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a proper and convex function, then $\xi$ is locally Lipschitz continuous in $\mathcal{D}^0(\xi)$.

Notice that from the previous definitions and from Rademacher’s theorem, it follows that for $\mathcal{L}^d$–a.e. $x$ in $\mathcal{D}^0(\xi)$ there exists $\nabla \xi(x)$ (and $\nabla \xi$ is locally bounded). If $\xi$ is differentiable in a point $x$, then we have $\xi(z) \geq \xi(x) + \nabla \xi(x) \cdot (z - x)$ for every $z \in \mathbb{R}^d$. From a geometrical point of view, this means that the graph of $\xi$ is always above its tangent hyperplane in $x$. Motivated by this, we define the subdifferential of a convex function as the collection of all the hyperplanes such that the graph of $\xi$ is above them.

Let $\mathcal{X}$ be a topological vector space with dual $\mathcal{X}^*$, $\xi : \mathcal{X} \to ]-\infty, +\infty]$ be convex and $x \in \mathcal{X}$. We define $\partial \xi(x)$ as the set $\partial \xi(x) := \{y \in \mathcal{X}^* : \xi(z) \geq \xi(x) + \langle y, z - x \rangle \}$ for every $z \in \mathcal{X}$. The set $\partial \xi(x)$ is called subdifferential of $\xi$ in $x$, and its elements are called subgradients. We also call subdifferential of $\xi$ the map $x \mapsto \partial \xi(x)$, seen as a function between $\mathcal{X}$ and $2^{\mathcal{X}^*}$. It follows immediately that, $\bar{x}$ is a minimum for $\xi$ if and only if $0 \in \partial \xi(\bar{x})$. If $\mathcal{X}$ is a topological vector space endowed with a norm $\| \cdot \|_{\mathcal{X}}$, then for all $x \in \mathcal{D}^0(\xi)$ there holds $\partial \xi(x) \neq \emptyset$. 
Notice that, this holds in any point, not almost everywhere! Nevertheless, it is not that useful if $D(\xi)$ has no interior points. In fact, it could be proved that $\partial \xi(x)$ is non–empty if $x \in D \subset D(\xi)$, for a suitable dense subset $D$, but such a proof would require much more sophisticated tools from convex analysis (e.g. approximated subdifferentials) which are far beyond the scope of these notes. We refer the reader to [86] or [100] for extensive treatments of the topic.

If a convex and proper function $\xi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is (Gateaux) differentiable in $x \in \mathcal{X}$, then $\partial \xi(x) = \{\nabla \xi(x)\}$. Conversely, if $\xi$ continuous in $\bar{x} \in \mathcal{X}$ and $\partial \xi(\bar{x}) = \{\bar{\varphi}\}$, then $\xi$ is (Gateaux) differentiable in $\bar{x} \in \mathcal{X}$ and $\nabla \xi(\bar{x}) = \bar{\varphi}$. Notice that, the continuity of $\xi$ in $x$ cannot be dropped in the reverse implication. Indeed, it is possible to have a convex set $A$ with empty interior such that in a point $\bar{x} \in A$ the normal cone reduces to $\{0\}$. Setting the indicator function of $A$ as the following map $I_A(x) := 0$ if $x \in A$ and $+\infty$ otherwise, we would have $\partial I_A(\bar{x}) = \{0\}$, but $I_A$ is not differentiable in $\bar{x}$. Notice that $I_A$ is convex if and only if $A$ is convex and $D(I) = A$. Here the key assumption is $A^o = D^o(I_A) = \emptyset$. Indeed, one could prove that $D^o(\xi) \neq \emptyset$ implies $\partial \xi(\bar{x})$ not a singleton whenever $\xi$ is discontinuous at $\bar{x}$ (i.e. at boundary points of the domain, the subdifferential is either empty or unbounded). In particular, if $D^o(\xi) \neq \emptyset$, then one has the nicer result which is: $\xi$ differentiable in $x$ if and only if $\partial \xi$ is a singleton. We refer to [86] for a set $A$ with the required properties and for a precise proof of the statement about subdifferentials in discontinuity points.

If $\mathcal{X}$ is a locally convex Hausdorff space, $\mathcal{X}^*$ is its dual space and $\xi : \mathcal{X} \to ]-\infty, +\infty]$ is a proper function, then, for any $y \in \mathcal{X}^*$, we define the Fenchel–Legendre conjugate of $\xi$ in $y$ as $\xi^* : \mathcal{X}^* \to ]-\infty, +\infty]$ given by $\xi^*(y) := \sup_{x \in \mathcal{X}} \{\langle y, x \rangle - \xi(x)\}$ and it is convex and lower semicontinuous. Moreover, if $\xi$ is convex, then $\xi^*$ is also proper. It follows directly that, for every $x \in \mathcal{X}$ and $y \in \mathcal{X}^*$, $\langle y, x \rangle \leq \xi(x) + \xi^*(y)$. Then, one can conclude that $\xi \leq \zeta$ implies $\xi^* \leq \zeta^*$, $\xi^{**} \leq \xi$, and $\xi^{***} = \xi^*$. If $\mathcal{X}$ is a normed space and $\xi : \mathcal{X} \to ]-\infty, +\infty]$ be a proper map, then the following facts are equivalent: $\xi$ is convex and lower semicontinuous, $\xi \equiv \zeta^*$ for some proper function $\zeta$, $\xi^{**} \equiv \xi$. Since the proof of these facts is quite long and requires a few concepts we do not need otherwise, we refer the reader to Proposition 2.5 in [115], which is known as the Legendre duality for lower semi-continuous convex function in the case $\mathcal{X} = \mathbb{R}^d$. If $\xi : \mathcal{X} \to ]-\infty, +\infty]$ is a convex, lower semicontinuous and proper map.
defined on the normed space $\mathcal{X}$, then, for every $x \in \mathcal{X}$ and $y \in \mathcal{X}^*$, there holds $y \in \partial\xi(x)$ if and only if $\langle y, x \rangle = \xi(x) + \xi^*(y)$ if and only if $x \in \partial\xi^*(y)$. Notice that, even if here we do not enter in the details, there is a sort of dual correspondence between strict convexity of $\xi$ and differentiability (or smoothness) of $\xi^*$, as Figure 2.3 in [115] tries to underline the loss of strict convexity in $\xi$ results in loss of regularity in $\xi^*$.

### B.2 Monotone functions in $\mathbb{R}^d$

Let $u : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ be a multifunction, i.e. $u$ maps points $x \in \mathbb{R}^d$ to subsets $u(x) \subset \mathbb{R}^d$. Then, for any fixed $B \subset \mathbb{R}^d$, $u$ is univalued on $B$ if $u(x)$ consists of at most one point for every $x \in B$. We define the domain of $u$ as $\text{Dom}(u) := \{x \in \mathbb{R}^d : u(x) \neq \emptyset\}$. We define the graph of $u$ as $\text{Graph}(u) := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : y \in u(x)\}$. We say $u$ is monotone if $\langle y_1 - y_2, x_1 - x_2 \rangle \geq 0$ for every $x_i \in \mathbb{R}^d$ and $y_i \in u(x_i) \ (i = 1, 2)$. For any given multifunctions $u, v$ we say that $u \supset v$, if $\text{Graph}(v) \subset \text{Graph}(u)$, i.e. if $v(x) \subset u(x)$ for every $x \in \mathbb{R}^d$. We say $u$ is maximal monotone if $v \supset u$ and $v$ is monotone implies $v = u$. In other words, $u$ is maximal monotone when it is maximal with respect to the inclusion among the monotone functions.

For univalued functions in $\mathbb{R}$, this monotone definition reduces to the usual monotonicity. Indeed, it is equivalent to $x_1 \leq x_2$ and $y_1 \leq y_2$ or $x_1 \geq x_2$ and $y_1 \geq y_2$. Therefore, any monotone nondecreasing map in $\mathbb{R}$ is a monotone function in the sense of the above definition.

An example of univalued monotone function on $\mathbb{R}^d$ is provided by the gradient $\nabla \xi$ of a convex function. Whenever $\xi$ is differentiable at $x, y \in \mathbb{R}^d$ then $(\nabla \xi(y) - \nabla \xi(x)) \cdot (y - x) \geq 0$. Indeed, for all $x, y$ in which $\xi$ is differentiable there holds $\xi(y) \geq \xi(x) + \nabla \xi(x) \cdot (y - x)$ and $\xi(x) \geq \xi(y) + \nabla \xi(y) \cdot (x - y)$. It is now sufficient to sum up the two inequalities and cancel the terms with $\xi$ to obtain the first inequality. Clearly, the same argument holds also for the subdifferential, whose definition was indeed modeled on this property. Anyway, we will come back later on the subdifferential and we will show that it is in fact maximal monotone.

Let $u$ be monotone. Then, $u$ is maximal monotone if and only if $\langle y - \alpha, x - \beta \rangle \geq 0$ for every $(x, y) \in \text{Graph}(u)$ implies $(\alpha, \beta) \in \text{Graph}(u)$. Indeed, by contradiction if this does not hold for a fixed $(\alpha, \beta)$, then we could define $v(x) := u(\alpha) \cup \{\beta\}$ if $x = \alpha$ and $u(x)$ otherwise. By properties of $u$ and the above statement, $v$ would be monotone and $v \supset u$, otherwise.
but $\alpha \in v(\beta) \setminus u(\beta)$, violating maximal monotonicity of $u$. On the other hand, take any $v : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ monotone such that $v \supset u$ and fix $(\alpha, \beta) \in \text{Graph}(v)$. By monotonicity, we would have $\langle y - \alpha, x - \beta \rangle \geq 0$ for every $(x, y) \in \text{Graph}(v)$. In particular, the same holds for all $(x, y) \in \text{Graph}(u) \subset \text{Graph}(v)$ and, by the statement above, $(\alpha, \beta) \in \text{Graph}(u)$. Therefore, $	ext{Graph}(v) \subset \text{Graph}(u)$, i.e. also $u \supset v$ and $u$ is maximal. We refer to [2] for a more complete development of the theory of maximal monotone functions and their properties.

Here we only state a few results and their application to the multifunction $\partial \phi$ for a convex map $\phi$. If $u : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ is maximal monotone, then $	ext{Graph}(u)$ is closed and $u(x)$ is a convex and closed (possibly empty) set for all $x \in \mathbb{R}^d$. A remarkable property of maximal monotone functions is that there exists a one-to-one correspondence between graphs of maximal monotone functions and graphs of 1-Lipschitz continuous functions.

Define the Cayley transformation $\Psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ as follows: $\Psi(x, y) := (x + y, -x + y)/\sqrt{2}$. If $u$ is a maximal monotone function, then $(u + \text{Id})^{-1}$ is defined on the whole $\mathbb{R}^d$ and $\Psi(\text{Graph}(u))$ is the graph of a 1-Lipschitz continuous function $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ (namely, $z \mapsto \varphi(z) := z - \sqrt{2}(u + \text{Id})^{-1}(\sqrt{2}z)$, which can be proved to be univalued and 1-Lipschitz, but we do not need this explicit form in the following). Moreover, if $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ is a 1-Lipschitz continuous function, then $\Psi^{-1}(\text{Graph}(\varphi))$ is the graph of a maximal monotone function on $\mathbb{R}^d$. In particular, $u$ is maximal if and only if $(u + \text{Id})$ is surjective.

Another important property of maximal monotone functions is that the set of points $x$ such that $u(x)$ is not univalued, is not too large. If $u$ is a maximal monotone function and set $\Xi_1(u) := \{x \in \mathbb{R}^d : \dim(u(x)) \geq 1\}$, where $\dim(u(x))$, is the Hausdorff dimension of $u(x) \in \mathbb{R}^d$, then $\dim(\Xi_1) \leq d - 1$, i.e. the Hausdorff dimension of $\Xi_1$ is at most $d - 1$. In particular, $\mathcal{H}^{d-1}(\Xi_1) < +\infty$. We do not give here a proof of the these previous results. You can find them in [2]. More precisely: the first statement corresponds to Proposition 1.1 in [2]; and the second is a particular case of Theorem 2.2 in [2] for $k = 1$. If $\xi$ is a proper convex function, then $\partial \xi$ is a maximal monotone function. Combining the results of the two previous paragraphs, one immediately obtains the following results. Let us first define the graph of the subdifferential of a convex function as $\text{Graph}(\partial \xi) := \{(x, y) \in \mathcal{X} \times \mathcal{Y} : y \in \partial \xi\} = \{(x, p) : \xi(x) + \xi^*(y) = \langle x, y \rangle\}$. If $\xi$ is a proper convex function, then $\text{Graph}(\partial \xi)$ is a closed set. Moreover, the set in which $\xi$ is not differentiable has finite $\mathcal{H}^{d-1}$ measure.
We finish this section with a last concept about convex functions, that we will then translate into the framework of c-concave functions later on. We already know that, in the case when \( \mathcal{X} = \mathcal{Y} = \mathbb{R} \), the graph is monotone in the sense that \((x, y) \in \text{Graph}(\partial \xi)\) if and only if \(\xi(x) + \xi^*(y) = x \cdot y\), for \(i = 1, 2\) implies \((y_2 - y_1) \cdot (x_2 - x_1) \geq 0\). Yet, not all monotone graphs are the graphs of the subdifferential of a convex functions, nor they are contained in one of such graphs. Hence, monotonicity is not enough to characterize gradients and subdifferential of convex functions (to be more precise, gradient vector functions that are monotone are gradient of convex functions, but monotonicity alone does not imply gradientness). A stronger notion is that of cyclical monotonicity: for any \(k \in \mathbb{N}\), every permutation \(\sigma\) and every finite family of points \((x_1, y_1), \ldots, (x_k, y_k) \in A \subset \mathbb{R}^d \times \mathbb{R}^d\), we have \(\sum_{i=1}^{k} x_i \cdot y_i \leq \sum_{i=1}^{k} x_i \cdot y_{\sigma(i)}\).

The word “cyclical” refers to the fact that, since every \(\sigma\) is the disjoint composition of cycles, it is enough to check this property for cyclical permutations, i.e. replacing \(\sum_{i=1}^{n} c(x_i, y_{\sigma(i)})\) with \(\sum_{i=1}^{n} c(x_i, y_{i+1})\) in the definition (with the obvious convention \(y_{n+1} = y_1\)). Notice that, if we take \(k = 2\) we get the usual definition of monotonicity, since \(x_1 \cdot y_1 + x_2 \cdot y_2 \geq x_1 \cdot y_2 + x_2 \cdot y_1\) if and only if \((y_1 - y_2) \cdot (x_1 - x_2) \geq 0\).

### B.3 Cyclical monotonicity, \(c\)-convexity and \(c\)-concavity

We start from the translation to the \(c\)-concave case of the definition of cyclical monotonicity. Let \(\mathcal{X}, \mathcal{Y}\) be arbitrary sets, and \(c : \mathcal{X} \times \mathcal{Y} \to ]-\infty, +\infty]\) be a function. A subset \(\Gamma \subset \mathcal{X} \times \mathcal{Y}\) is said to be \(c\)-cyclically monotone if, for any \(k \in \mathbb{N}\), and any family \((x_1, y_1), \ldots, (x_k, y_k) \in \Gamma\), holds the inequality \(\sum_{i=1}^{k} c(x_i, y_i) \leq \sum_{i=1}^{k} c(x_i, y_{i+1})\) with the convention \(y_{k+1} = y_1\). A \(c\)-cyclically monotone set \(\Gamma\) is said to be maximal if it is maximal with respect to the set inclusion. A transference plan is said to be \(c\)-cyclically monotone if it is concentrated on a \(c\)-cyclically monotone set. In particular, in the case \(\mathcal{X} = \mathcal{Y} = \mathbb{R}^d\) and \(c(x, y) = |x - y|^2\), a set \(\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d\) is \(c\)-cyclically monotone if for all \(k \in \mathbb{N}\) and for all families \((x_1, y_1), \ldots, (x_k, y_k) \in \Gamma\) there holds \(\sum_{i=1}^{m} |x_i - y_i|^2 \leq \sum_{i=1}^{m} |x_i - y_{i+1}|^2\), with the convention \(y_{m+1} = y_1\). Equivalently, we could require \(\sum_{i=1}^{m} y_i \cdot (x_{i+1} - x_i) \leq 0\), with the convention \(x_{m+1} = x_1\).

Let start with two trivial examples in \(\mathcal{X} = \mathcal{Y} = \mathbb{R}^2\) and \(c(x, y) = |x - y|\), just to learn...
how to handle the definition above: \( \Gamma_1 := \{(0,0), (1,1), (2,2)\} \) is a \( c \)-cyclically monotone set while \( \Gamma_2 := \{(0,0), (1,0), (0,1), (1,1)\} \) is not; indeed, with \( k = 2 \) and \( (1,0), (0,1) \), it fails to satisfy the above definition. Note that a power \(|x - y|^p, p > 1\), would work as well, but this is not really an interesting example anyway. A famous theorem by Rockafellar [100], states that every cyclically monotone set is contained in the graph of the subdifferential of a convex function. We will not prove this theorem here, both since we do not really need it, and since we will see it as a particular case of a theorem on \( c \)-concave functions.

**Theorem B.1 (Rockafellar’s Theorem)** A nonempty set \( \Gamma \subset \mathbb{R}^n \times \mathbb{R}^n \) is cyclically monotone (with respect to \( c(x, y) = |x - y|^2 \)) if and only if it is included in the graph of the subdifferential of a proper, lower semicontinuous and convex function \( \xi : \mathbb{R}^n \to (-\infty, +\infty] \). Moreover, every maximal cyclically monotone set coincides with such a graph.

Let \( \mathcal{X}, \mathcal{Y} \) be sets and \( c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\} \) a function. A function \( \xi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\} \) is said to be a proper \( c \)-convex function if it is not identically \(+\infty\), and there exists \( \psi : \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\} \) such that \( \xi(x) = \sup_{y \in \mathcal{Y}} (\psi(y) - c(x, y)) \) for every \( x \in \mathcal{X} \), with the convention that \((+\infty) - (+\infty) = -\infty\). A function \( \zeta : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\} \) is said to be a proper \( c \)-concave function if it is not identically \(-\infty\), and there exists \( \varphi : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\} \) such that \( \zeta(y) = \inf_{x \in \mathcal{X}} (\varphi(x) + c(x, y)) \) for every \( y \in \mathcal{Y} \), with the convention that \((-\infty) + (+\infty) = +\infty\). Note that, roughly speaking, a \( c \)-convex function is a function whose graph can be “caressed” from a graph with an object having the shape of \(-c\) (this shape might of course vary when you change point). Figure 5.2. in [116] could help to visualize the situation. If \( c(x, y) = -x \cdot y \) on \( \mathbb{R}^d \times \mathbb{R}^d \), then the \( c \)-convexity coincides with the usual convexity plus lower semicontinuity (indeed, we would have \( \xi \) \( c \)-convex if and only if \( \xi = (-\varphi)^* \) for some \( \varphi \) and the equivalence would follow from results of the preceding section).

Let \( \mathcal{X}, \mathcal{Y} \) be sets and \( c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\} \) a function and \( \xi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\} \) \( c \)-convex. The \( c \)-transform of \( \xi \) is the function \( \xi_c : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\} \) defined by \( \xi_c(y) = \inf_{x \in \mathcal{X}} (\xi(x) + c(x, y)) \) for every \( y \in \mathcal{Y} \). The \( c \)-subdifferential of \( \xi \) is the is the \( c \)-cyclically monotone set defined by \( \partial_c \xi := \{(x, y) \in \mathcal{X} \times \mathcal{Y} : \xi_c(y) - \xi(x) = c(x, y)\} \). The functions \( \xi \) and \( \xi_c \) are said to be \( c \)-conjugate. The \( c \)-subdifferential of \( \xi \) at point \( x \in \mathcal{X} \) is \( \partial_c \xi(x) := \{y \in \mathcal{Y} : (x, y) \in \partial_c \xi\} \), or equivalently, the set of \( y \in \mathcal{Y} \) such that for all \( \bar{x} \in \mathcal{X} \) there
holds $\xi(x) + c(x, y) \leq \xi(\bar{x}) + c(\bar{x}, y)$. Notice that, indeed, the equivalently came from that $(x, y) \in \partial_\xi \xi$ if and only if $\xi(x) + c(x, y) = \xi_c(y) = \inf_{\bar{x} \in \mathcal{X}} (\xi(\bar{x}) + c(\bar{x}, y))$. It is immediate to notice that $\xi_c$ is $c$-concave, whenever $\xi$ is $c$-convex (Even more, the definition of $c$-concave functions can be restated as $\zeta = \psi^c$ for some $\psi$). This behavior is different from the usual one with convex functions and their Fenchel-Legendre transforms. Namely, if $c(x, y) = -x \cdot y$ and $\xi$ convex, then one always has $\xi_c = -\xi^*$. 

Let $\mathcal{X}, \mathcal{Y}$ be sets and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ a function and $\zeta : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be $c$-concave. The $c$-transform of $\zeta$ is the function $\zeta^c : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ defined by $\zeta^c(x) = \sup_{y \in \mathcal{Y}} (\zeta(y) - c(x, y))$ for every $x \in \mathcal{X}$. The $c$-superdifferential of $\zeta$ is the is the $c$-cyclically monotone set defined by $\partial^c \zeta := \{(x, y) \in \mathcal{X} \times \mathcal{Y} : \zeta(y) - \zeta^c(x) = c(x, y)\}$. The functions $\zeta$ and $\zeta^c$ are said to be $c$-conjugate. The $c$-superdifferential of $\zeta$ at point $y \in \mathcal{Y}$ is $\partial^c \zeta(y) := \{x \in \mathcal{X} : (x, y) \in \partial^c \zeta\}$, or equivalently, the set of $x \in \mathcal{X}$ such that for all $\tilde{y} \in \mathcal{Y}$ there holds $\zeta(y) - c(x, y) \geq \zeta(\tilde{y}) - c(x, \tilde{y})$. As in the previous definition, notice that the equivalently came also from that $(x, y) \in \partial^c \zeta$ if and only if $\zeta(y) - c(x, y) = \zeta^c(x) = \sup_{y \in \mathcal{Y}} (\zeta(y) - c(x, y))$. Notice also that both the $c$-subdifferentials $\partial_c \xi$ and the $c$-superdifferentials $\partial^c \zeta$ are $c$-cyclically monotone sets.

Now, we want to stress that, in all definitions above, we never assumed any measurability and $c$ is any function defined on the product $\mathcal{X} \times \mathcal{Y}$. Hence, a priori, measurability of $\xi, \zeta$ or $\xi_c, \zeta^c$ is not guaranteed. Assuming $c$ continuous, we would be able to say more; for instance, a $c$-convex (resp. $c$-concave) function is lower semicontinuous (resp. upper semicontinuous) and its $c$-subdifferential (resp. $c$-superdifferential) is a closed set. Despite of the loss of symmetry between $\mathcal{X}$ and $\mathcal{Y}$ induced by using infimum or supremum depending on the $c$-convexity or $c$-concavity, and of the differences with standard convexity (as underlined in the previous paragraph), previous definitions allow to recover some of the properties of Fenchel-Legendre transform for convex functions, as the following properties show. Before that let us begin by recalling the $c$-convexification of $\xi$ (resp. the $c$-concaveification of $\zeta$) to be defined by $\xi_{cc} = (\xi_c)_c$ (resp. $\zeta_{cc} = (\zeta^c)_c$). More explicitly, $\xi_{cc}(x) = \sup_{y \in \mathcal{Y}} \inf_{\bar{x} \in \mathcal{X}} (\xi(\bar{x}) + c(\bar{x}, y) - c(x, y))$ and $\zeta_{cc}^c(y) = \inf_{\bar{x} \in \mathcal{X}} \sup_{y \in \mathcal{Y}} (\zeta(y) - c(x, \tilde{y}) + c(x, y))$. If $\mathcal{X}, \mathcal{Y}$ are sets and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}, \xi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ and $\zeta : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ are functions, then $\xi_{ccc} := ((\xi_{cc})_c)_c = \xi_c$ and $\zeta_{ccc}^c := ((\zeta_{cc})^c)_c = \zeta^c$. The following proposition may be taken as the main justification for
the concept of \(c\)-convexity and \(c\)-concavity and which gives an alternative characterizations of them.

**Proposition B.2** Let \(\mathcal{X}, \mathcal{Y}\) be sets and \(c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}\), \(\xi : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}\) and \(\zeta : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}\) be functions. Then, \(\xi\) is \(c\)-convex if and only \(\xi = \xi_{cc}\). Moreover, \(\zeta\) is \(c\)-concave if and only \(\zeta = \zeta_{cc}\).

**B.4 Differentiability in the \(p^{\text{th}}\)-Wasserstein space**

Let \(U : \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R} \cup \{\pm \infty\}\), we denote by \(\partial U(\mu)\) the subdifferential of \(U\) at \(\mu\), as introduced in [20] (see Definition 5.3). As shown in [20], for \(\lambda\)-convex functionals, this definition coincides with the one in (10.3.12) of [17]. Since \(\|\cdot\|_{L^p(\mu)}\) is uniform and \(\partial U(\mu)\) is a closed convex subset of \(L^p(\mu)\), it admits a unique element of minimal norm. As it is customary in subdifferential analysis, we denote that element by \(\nabla_\mu H(\mu)\). We refer to it as the gradient of \(U\) with respect to the Wasserstein distance \(W_p\). The super differential of \(U\) at \(\mu\) is denoted by \(\partial U(\mu)\) and consists of the \(\xi\) such that \(-\xi\) belongs to \(\partial(-U)(\mu)\). We say that \(U\) is differentiable at \(\mu\) if \(\partial U(\mu)\) and \(\partial U(\mu)\) are nonempty. In that case, both sets coincide.

For any \(1 < p < +\infty\) and \(\mu \in \mathcal{P}_p(\mathbb{R}^d)\), the tangent space to \(\mathcal{P}_p(\mathbb{R}^d)\) can be defined by the closure of the set \(\{\partial \xi : \xi \in C^\infty_c(\mathbb{R})\}\) in \(L^p(\mu)\), i.e. it denoted by \(\mathcal{T}_\mu \mathcal{P}_p(\mathbb{R}^d) := \{\partial \xi : \xi \in C^\infty_c(\mathbb{R})\}^{L^p(\mu)}\). We refer to \(\mathcal{T}_\mu \mathcal{P}_p(\mathbb{R}^d)\) as the tangent space to \(\mathcal{P}_p(\mathbb{R}^d)\) at \(\mu\) (see section 8.5 of [17]). In this case it is easy to check that \(\mathcal{T}_\mu \mathcal{P}_p(\mathbb{R}^d) = L^p(\mu)\) for some \(1 < p < +\infty\).
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Biography

Mohamed H. Amsaad was born in Benghazi, Libya, in 1978. He graduated from high school in 1996. He earned his Bachelors of Sciences and “first” Masters of Science degrees in Mathematics in June of 2002 and 2006, respectively, from the University of Benghazi in Libya. After he spent almost four years for teaching as mathematics teacher in High schools, he worked as Assistant Lecturer at the Department of Mathematics and Computer Science at Qaminis College of Education in which he also served as a Department Chair. In January 2010, Mr. Amsaad joined the Department of Mathematics at West Virginia University in the USA as a Graduate Student. He was supported by Libyan Ministry of Higher Education and by the State of West Virginia and granted assistantship in different occasions on both positions of Teaching Assistant and Research Assistant. He received his “second” Master’s and Doctorate of Philosophy degrees in Applied Mathematics from West Virginia University, in May 2015 and May 2016 respectively.