Scuola Normale Superiore of Pisa

and

École Normale Supérieure of Lyon

Phd thesis

24th October 2007

Optimal transportation and action-minimizing measures

Alessio Figalli

a.figalli@sns.it

Advisor
Prof. Luigi Ambrosio
Scuola Normale Superiore of Pisa

Advisor
Prof. Cédric Villani
École Normale Supérieure of Lyon.

Contents

Introduction

The Monge transportation problem is more than 200 years old [110], and it has generated in the last years a huge amount of work.

Originally Monge wanted to move, in 3-space, a rubble (déblais) to build up a mound or fortification (remblais) minimizing the cost. Now, if the rubble consists of masses, say $m_1, \dots, m_n$ at locations $\{x_1, \dots x_n\}$, one should move them into another set of positions $\{y_1, \dots, y_n\}$ by minimizing the weighted traveled distance. Therefore one should try to minimize

$$\sum_{i=1}^{n} m_i |x_i - T(x_i)|,$$

over all bijections $T : \{x_1, \dots x_n\} \to \{y_1, \dots, y_n\}$, where d is the usual Euclidean distance on 3-space.

Nowadays, one would be more interested in minimizing the energy cost rather than the traveled distance. Therefore one would try rather to minimize

$$\sum_{i=1}^{n} m_i |x_i - T(x_i)|^2.$$

Of course, it is desirable to generalize this to continuous, rather than just discrete, distributions of matter. To this aim, Monge transportation problem can be stated in the following general form: given two probability measures μ and ν, defined on the measurable spaces X and Y, find a measurable map $T : X \to Y$ with

$$T_\sharp \mu = \nu,$$

i.e.

$$\nu(A) = \mu\left(T^{-1}(A)\right) \quad \forall A \subset Y \text{ measurable},$$

and in such a way that T minimizes the transportation cost. This last condition means

$$\int_X c(x, T(x))\, d\mu(x) = \min_{S_\sharp \mu = \nu} \left\{ \int_X c(x, S(x))\, d\mu(x) \right\},$$

where $c : X \times Y \to \mathbb{R}$ is some given cost function, and the minimum is taken over all measurable maps $S : X \to Y$ with $S_\sharp \mu = \nu$. When the transport condition $T_\sharp \mu = \nu$ is satisfied, we say that T is a *transport map*, and if T minimizes also the cost we call it an *optimal transport map*.

In the development of the theory of optimal transportation, as well as in the development of other theories, it is important on the one hand to explore new variants of the original problem, on the other hand to figure out, in this emerging variety of problems, some common (and sometimes unexpected) features. This kind of analysis is the main scope of our thesis.

The problems we will consider are:

1. The optimal transportation problem on manifolds with geometric costs: we study the problem of the existence and the uniqueness of an optimal transport map on arbitrary manifolds (without any condition on the sectional curvature), for costs of the form $c(x, y) := \inf_\gamma \int_0^1 L(\gamma, \dot\gamma)\, dt$, where the infimum is among all absolutely continuous curves from x to y, and $L(x, v)$ is a Tonelli Lagrangian.

2. The optimal irrigation problem: this is a generalization of the classical optimal transportation problem, where one wants to connect a source measure to a target measure using a "transport structure" in such a way that the mass moves, as much as possible, in a grouped way. The motivation for such a problem comes from the modelization of many biological and engineering structures.

3. The Brenier variational theory of incompressible flows: starting from the geometrical interpretation of the Euler equations for incompressible fluids as a geodesic equation in the space of the measure-preserving diffeomorphisms, one can look for solutions of the Euler equations by minimizing the action functional. This leads to the introduction of relaxed models and their study from a calculus of variation point of view.

4. The Aubry-Mather theory and the solutions of Hamilton-Jacobi equations: the regularity and the uniqueness of viscosity solutions of the Hamilton-Jacobi equation is linked to the smallness of certain sets appearing in Lagrangian dynamics. Thus we will be interested to estimate the Hausdorff dimension of the so called quotient Aubry set.

5. The DiPerna-Lions theory for martingale solutions of stochastic differential equations: this theory allows in some sense to prove a sort of existence and uniqueness for an ordinary differential equation for almost every initial condition once one has some well-posedness result at the level of the associated transport equation. Our

aim will be to develop such a theory in the case of an ordinary differential equation perturbed by an irregular noise.

We remark that the first three topics are all variants of the optimal transportation problem. Moreover, even though the last topic is only loosely related to optimal transportation, at the technical level many connections arise, and the study of all them reveals some new connections. For instance, Bernard and Buffoni [22] have recently shown how one can fit Mather's theory, as well as optimal transportation problems on manifolds with a geometric cost, in the framework of measures in the space of action-minimizing curves. We proceed in this research of a general unified framework, proving that also variational solutions of the Euler equations [8] can be seen in this perspective, with a possibly non-smooth action induced by the pressure field. Also the last two topics present some links with the first three. For instance, the proof in [23] on the existence and uniqueness of an optimal transport plan strongly relies on the regularity properties of solutions of Hamilton-Jacobi equations, while the natural framework which allows to develop a theory à la DiPerna-Lions for martingale solutions of stochastic differential equations turns out to be the one of the measures in the space of paths, which, as we said, is natural also in the optimal transportation problem and in the variational study of the Euler equations.

Let us give a quick overview on all these subjects (each chapter contains a more detailed mathematical and bibliographical description of the single problems), providing also an outline of thesis' content. All the results in this thesis have been presented in a series of papers (accepted, submitted, or in preparation), originating from several collaborations developed during the PhD studies.

1. As we explained above, one is interested to find a transport map $T : X \to Y$ from μ to ν which minimizes the transportation cost, that is

$$\int_X c(x, T(x))\, d\mu(x) = \min_{S_\sharp \mu = \nu} \left\{ \int_X c(x, S(x))\, d\mu(x) \right\}.$$

Even in Euclidean spaces, with the cost c equal to the Euclidean distance or its square, the problem of the existence of an optimal transport map is far from being trivial. Moreover, it is easy to build examples where the Monge problem is ill-posed simply because there is no transport map: this happens for instance when μ is a Dirac mass while ν is not. This means that one needs some restrictions on the measures μ and ν.

The major advance on this problem is due to Kantorovitch, who proposed in [85], [86] a notion of weak solution of the optimal transport problem. He suggested to look for *plans* instead of transport maps, that is probability measures γ in $X \times Y$ whose marginals are μ and ν, i.e.

$$(\pi_X)_\sharp \gamma = \mu \quad \text{and} \quad (\pi_Y)_\sharp \gamma = \nu,$$

where $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are the canonical projections. Denoting by $\Pi(\mu, \nu)$ the set of plans, the new minimization problem becomes then the following:

$$C(\mu, \nu) = \min_{\gamma \in \Pi(\mu,\nu)} \left\{ \int_{M \times M} c(x, y)\, d\gamma(x, y) \right\}. \tag{0.0.1}$$

If γ is a minimizer for the Kantorovich formulation, we say that it is an *optimal plan*. Due to the linearity of the constraint $\gamma \in \Pi(\mu, \nu)$, it turns out that weak topologies can be used to provide existence of solutions to (0.0.1): this happens for instance whenever X and Y are Polish spaces and c is lower semicontinuous (see [118], [132, Proposition 2.1] or [133]). The connection between the formulation of Kantorovich and that of Monge can be seen by noticing that any transport map T induces the plan defined by $(\mathrm{Id}_X \tilde{\times} T)_\sharp \mu$ which is concentrated on the graph of T, where the map $\mathrm{Id}_X \tilde{\times} T : X \to X \times Y$ is defined by

$$\mathrm{Id}_X \tilde{\times} T(x) = (x, T(x)).$$

Thus, the problem of showing existence of optimal transport maps reduces to prove 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. The first existence and uniqueness result is due to Brenier. In [30] he considers the case $X = Y = \mathbb{R}^n$, $c(x, y) = |x - y|^2$, and he shows that, if μ is absolutely continuous with respect to the Lebesgue measure, there exists a unique optimal transport map. After this result, many researchers started to work on the problem, showing existence of optimal maps with more general costs, both in a Euclidean setting (for example Caffarelli, Evans, Ambrosio and Pratelli, Trudinger and Wang, McCann, Feldman), in the case of compact manifolds (McCann, Bernard and Buffoni), and in some particular classes on non-compact manifolds (Feldman and McCann), see Section 1.1 for precise references.

A fact which is now well understood is that the choice of the cost changes completely the structure of the problem. In particular, completely different are the cases $c(x, y) = |x - y|^p$, with $p > 1$, with respect to the case $c(x, y) = |x - y|$ (the latter has been solved in the Euclidean case many years after the result of Brenier: first by Evans and Gangbo [60] under some regularity assumptions on the measures, then by Caffarelli, Feldman and McCann [39] under much weaker assumptions on the measures, and finally, in a more general case, by Ambrosio and Pratelli [13]). Indeed, the strict convexity of $|x - y|^p$ for $p > 1$ allows to prove existence and uniqueness of the transport map under the absolute continuity assumption on μ. On the other hand, in the case $c(x, y) = |x - y|$ one can still prove existence of optimal maps if μ is absolutely continuous, but no uniqueness result can be expected. This is a consequence, even on the real line, of the so-called *book-shifting* phenomenon: taking $\mu = \mathscr{L}^1 \lfloor_{[0,1]}$ and $\nu = \mathscr{L}^1 \lfloor_{[1/2,3/2]}$, the two maps $T_1(x) = x + 1/2$ and $T_2(x) = (x + 1)\chi_{[0,1/2]}(x) + x\chi_{[1/2,1]}(x)$ are both optimal. This is a special case of the general fact that, with the cost $c(x,y)=|x-y|$, one can find an optimal transport map imposing also that the common mass between μ and ν stays fixed. In Chapter 1, following a joint work with Albert Fathi [66], we show existence and uniqueness of an optimal transport map in a very general setting, which includes the case of "geometric" costs on manifolds, that is costs given by

$$c(x,y) := \inf_{\gamma(0)=x,\gamma(1)=y} \int_0^1 L(\gamma(t),\dot\gamma(t))\,dt,$$

where $L : TM \to \mathbb{R}$ is a Tonelli Lagrangian. This is the most general known result, since it is valid for a wide class of cost functions and it does not require any global assumption on the manifold (say, a bound on the sectional curvature). To this aim, we will need to understand the regularity of the cost alongs extremals, a problem which is closely linked to weak KAM theory and the regularity of solutions of the Hamilton-Jacobi equation, see also Section 6.2.2.

Moreover, in Section 1.5 we will study the so-called "displacement interpolation", which is a way to connect measures using the optimal transportation. For instance, suppose that μ_0 and μ_1 are two absolutely continuous measures in $\mathbb{R}^d$, and let $T : \mathbb{R}^d \to \mathbb{R}^d$ be the optimal transport map from μ_0 to μ_1 (as we said above, existence and uniqueness of the optimal transport map in this special case is due to Brenier [30]). Then, instead of "connecting" μ_0 to μ_1 in a linear way (that is $\mu_t = (1-t)\mu_0 + t\mu_1$), one can consider the interpolation $\mu_t := ((1-t)\,\mathrm{Id} + tT)_\#\mu_0$, which is called "displacement interpolation". An interesting feature is that, from the convexity of certain funtionals along such curves, one can deduce existence, uniqueness and stability of the gradient flows of such functionals obtaining many interesting properties for Fokker-Planck-type evolution equations such as the porous medium equation (see [42], and see also [11] for an introduction and a wide bibliography on this subject).

The convexity of certain suitable functionals on Riemannian manifolds allows to express Ricci curvature bounds on the manifold. In Section 1.7, following a joint work with Cedric Villani [78], we use the general results on optimal transport maps mentionned above to study the link between more possible notions of "displacement convexity" (i.e. convexity along displacement interpolations) and to prove their equivalence.

Finally, in Section 1.8 we will generalize the existence and uniqueness of the optimal transport map without assumptions on the finiteness of the transportation cost, and we will also prove that the optimal transport map on a general manifold is approximately differentiable a.e. whenever the cost is given by $c(x,y) = d^2(x,y)$ [75].

2. Another kind of transport problem is the so-called *irrigation problem*. Starting from the observation of the frequent occurrence of branched networks in nature (plants and trees, river basins, bronchial and cardiovascular systems) and in man designed structures (communication networks, electric power supply, water distribution

or drainage networks), and observing that the common function of such networks is to transport some goods from an initial distribution (the supply) to another one (the demand), we are interested in finding models which describe such fenomena. This was done in [82, 98, 135, 25, 24, 29] by considering cost functions that encode the efficiency of a transport induced by some structure. Branched structures, as the ones observed in nature, then arise as the optimal structures along which the transport takes place.

The first model is due to Gilbert [82]: given two atomic probability measures $\mu = \sum_{i=1}^{N_1} a_i\delta_{x_i}$ and $\nu = \sum_{i=1}^{N_2} b_i\delta_{y_i}$, find a finite, oriented and weighted graph $\Gamma = (v_h, p_h)$ (v_h are the vectors that orient the graph, p_h the weights), which satisfies Kirchhoff's law at the junctures (the mass which enters is equal the mass which exits, except at the points x_i where a mass a_i exits, and at the points y_j where a mass b_j enters). One then looks for a graph which minimizes the transportation cost

$$C^\alpha(\Gamma) = \sum_h |v_h| p_h^\alpha$$

with $0 \le \alpha \le 1$. The motivation for introducing such a parameter α is that, since the function $t \mapsto t^\alpha$ is sub-additive for $0 \le \alpha \le 1$, the inequality $(p_{h_1} + p_{h_2})^\alpha \le p_{h_1}^\alpha + p_{h_2}^\alpha$ holds, and thus the mass has interest to concentrate and to move together. This problem has been recently generalized (by Xia, Morel, Bernot, Solimini, etc.) to the case of arbitrary probability measures, and one arrives at problems where the optimal objects have a branched structure, and the optimal transportation costs $C^\alpha(\mu, \nu)$ give rise to a distance which metrizes the weak convergence. In order to extend the above problem to arbitrary target and source measures, a "probabilistic" formalism that has been considered in [98, 25, 24] is the one of *traffic plans*, which are suitable probability measures in the space of continuous paths which "connect" two fixed measures μ and ν. In this framework, all particles are indexed by the set $\Omega := [0, 1]$, and to each $\omega \in \Omega$ is associated a 1-Lipschitz path $\chi(\omega, \cdot)$ in $\mathbb{R}^N$. This is a Lagrangian description of the dynamic of particles that can be encoded by the image measure $\mathbf{P}_\chi$ of the map $\omega \mapsto \chi(\omega, \cdot)$ (which is therefore a measure on the set of 1-Lipschitz paths). To each traffic plan one can associate a suitable cost function which has to incorporate the principle that it is more efficient to transport mass in a grouped way rather than in a separate way. Like in the discrete case considered by Gilbert, to embed this principle the costs incorporate a parameter $\alpha \in [0, 1]$ and make use of the concavity of $x \mapsto x^\alpha$. Once the cost and the measures μ and ν are given, one can consider what is called *the irrigation problem* by some authors [25, 24, 26], i.e. the problem of minimizing the cost among structures transporting μ to ν.

In Chapter 2 we study different kinds of possible costs, and in some case we show their equivalence. Moreover, we study the properties of the costs when seen as functions of the parameter α, and we use this analysis to show a stability property of minimizers.

This is a joint work with Marc Bernot [27].

3. The velocity of an incompressible fluid moving inside a region D is mathematically described by a time-dependent and divergence-free vector field $\boldsymbol{u}(t,x)$ which is parallel to the boundary ∂D. The *Euler equations* for incompressible fluids describes the evolution of such velocity field $\boldsymbol{u}$ in terms of the pressure field p:

$$\begin{cases} \partial_t \boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = -\nabla p & \text{in } [0,T]\times D, \\ \operatorname{div}\boldsymbol{u} = 0 & \text{in } [0,T]\times D, \\ \boldsymbol{u}\cdot n = 0 & \text{on } [0,T]\times \partial D. \end{cases} \tag{0.0.2}$$

Let us assume that $\boldsymbol{u}$ is smooth, so that it produces a unique flow g. Writing the Euler equations in terms of g, we get

$$\begin{cases} \ddot{g}(t,a) = -\nabla p\,(t, g(t,a)) & (t,a)\in[0,T]\times D, \\ g(0,a) = a & a\in D, \\ g(t,\cdot)\in \mathrm{SDiff}(D) & t\in[0,T], \end{cases} \tag{0.0.3}$$

where $\mathrm{SDiff}(D)$ denotes the space of measure-preserving diffeomorphisms of D. Viewing $\mathrm{SDiff}(D)$ as an infinite-dimensional manifold with the metric inherited from the embedding in L^2, and with tangent space made by the divergence-free vector fields, Arnold interpreted the equation above, and therefore (0.0.2), as a *geodesic* equation on $\mathrm{SDiff}(D)$ [15]. According to this intepretation, one can look for solutions of (0.0.3) by minimizing

$$T\int_0^T \int_D \frac{1}{2}|\dot{g}(t,x)|^2\, d\mu_D(x)\, dt$$

among all paths $g(t,\cdot):[0,T]\to \mathrm{SDiff}(D)$ with $g(0,\cdot)=f$ and $g(T,\cdot)=h$ prescribed. This minimization problem presents many difficulties from the calculus of variations point of view (mainly because of the incompressibility constraint), and also gives rise to many interesting questions. Brenier [31, 35] introduced two relaxed models to study this problem. In particular, in [31], he defined a *generalized incompressible flow* as a probability measure $\boldsymbol{\eta}$ on $\Omega(D) := C([0,T],D)$ such that

$$(e_t)_\#\boldsymbol{\eta} = \mathscr{L}^d \llcorner D \quad \forall t\in[0,T], \qquad (e_0,e_T)_\#\boldsymbol{\eta} = (id\times h)_\#\mathscr{L}^d\llcorner D,$$

and defined the *action* of $\boldsymbol{\eta}$ as

$$\mathscr{A}(\boldsymbol{\eta}) := \frac{1}{2}\int_{\Omega(D)}\int_0^T |\dot{\omega}(t)|^2\, dt\, d\boldsymbol{\eta}(\omega).$$

The existence of a minimizer can be proved by a standard compactness and semicontinuity argument. Moreover, by a duality argument, Brenier introduced the pressure field, and proved that a minimizer of the above variational problem $\boldsymbol{\eta}$ solves in a "weak" sense the Euler equations: for all $h \in C_c^\infty(0,1)$ and $\boldsymbol{w}$ smooth compactly supported vector field on D, we have

$$\int_{\Omega(D)} \int_0^T \dot{\omega}(t) \cdot \frac{d}{dt}[h(t)\boldsymbol{w}(\omega(t))]\, dt\, d\boldsymbol{\eta}(\omega) = \langle D_x p(t,x), h(t)\boldsymbol{w}(x)\rangle$$

in the distributional sense.
In particular, this condition identifies uniquely the pressure field p (as a distribution) up to trivial modifications, i.e. additive perturbations depending on time only. We also remark that, if the measure $\boldsymbol{\eta}$ is given by

$$\int_{\Omega(D)} f(\omega)\, d\boldsymbol{\eta}(\omega) = \int_D f(t \mapsto g(t,x))\, dx$$

with $g : [0,T] \to SDiff(D)$ smooth, then $\boldsymbol{u}(t,x) := \partial_t g(t, g^{-1}(t,x))$ is a solution of the Euler equations. Now an important problem is to study the structure of minimizers, finding necessary and sufficient conditions for optimality, a question which will be addressed and solved in Chapter 3 following a joint work with Luigi Ambrosio [8]. As we already said, the results we prove show a somehow unexpected connection between the variational theory of incompressible flows and the theory developed by Bernard-Buffoni [22] of measures in the space of action-minimizing curves.

Indeed, first we refine a little bit the deep analysis made in [35] of the regularity of the gradient of the pressure field: Brenier proved that the distributions $\partial_{x_i} p$ are locally finite measures in $(0,T) \times D$, but this information is not sufficient (due to a lack of time regularity) to imply that p is a function. In Section 3.7 we improve this result showing that $p \in L^2_{\mathrm{loc}}\big((0,T); BV_{\mathrm{loc}}(D)\big)$ (this has been done in another joint work with Luigi Ambrosio [9]). In particular p is a function at least in some $L^1_{\mathrm{loc}}(L^r_{\mathrm{loc}})$ space, for some $r > 1$. We can therefore develop a refined analysis of the necessary and sufficient optimality conditions for action-minimizing curves in $\Gamma(D)$ (see Section 3.6) which involve the Lagrangian

$$\mathcal{L}_p(\gamma) := \int \frac{1}{2}|\dot{\gamma}(t)|^2 - p(t,\gamma(t))\, dt,$$

the (locally) minimizing curves for $\mathcal{L}_p$ and the value function induced by $\mathcal{L}_p$.

We also remark that the possibility of deducing such regularity result for the pressure is based on the equivalence, proved in Section 3.4, of the above mentioned Brenier model and the Eulerian-Lagrangian model introduced by the same author in [35] (see Section 3.3.3). Indeed, the regularity of p is easier to study within the latter model.

4. As we said, an important connection between Mather's theory as well as optimal transportation problems on manifolds exists [22, 23]. The key point is that cost functions induced by Tonelli Lagrangians solve an Hamilton-Jacobi equation.

Important for studying the dynamic of a Lagrangian system and for having uniqueness of solutions of the Hamilton-Jacobi equation

$$H(x, d_x u) = c$$

is to understand the structure of some subsets of the tangent space which capture the properties of the dynamic. Mather [105] proposed as an important problem to show that the *quotient Aubry set* is totally disconnected if the Lagrangian (or, equivalently, the Hamiltonian) is smooth. In Chapter 4 this problem will be completely solved up to dimension 3, and in many particular cases in higher dimension, following a joint work with Albert Fathi and Ludovic Rifford [67].

To understand the key idea of the proof, let us consider the particular case

$$H(x, p) = \frac{1}{2}|p|^2 + V(x),$$

and without loss of generality let us assume $\max_x V(x) = 0$. Then in this case the Hamilton-Jacobi equation one is interested in becomes

$$\frac{1}{2}|d_x u|^2 + V(x) = 0$$

(the value $c = 0$ is the Mañé critical value for the above Hamiltonian), and the projected Aubry set is the set $\{V = 0\}$. As shown in Section 4.2, the key point to show that the *quotient Aubry set* is totally disconnected (or small in the sense of the Hausdorff dimension) is to prove a sort of Sard-type theorem for critical subsolution of the Hamilton-Jacobi equation (that is functions u which satisfy $\frac{1}{2}|d_x u|^2 + V(x) \leq 0$), showing that the image of the set $\{V = 0\}$ under the map $u : M \to \mathbb{R}$ has zero Lebesgue measure. Although the function u is only C^1, and so the classical Sard theorem cannot be applied, in this case one has the extra information

$$|d_x u|^2 \leq -2V(x),$$

and the function $V(x)$ is smooth by assumption. One can therefore use the regularity of V to deduce that u is really "flat" near $\{V = 0\}$, and so to deduce the Sard-type result.

5. In Chapter 5, we will develop a theory à la DiPerna-Lions for martingale solutions, in the sense of Stroock-Varadhan, of stochastic differential equations.

In [56, 4], the authors developed a theory which, roughly speaking, allows to prove existence and uniqueness in a weak sense for solutions of ordinary differential equations with nonsmooth coefficients. This theory is bsed on the classical links between the *transport* (or the *continuity*) *equation*

$$\partial_t \mu + \operatorname{div}(b\mu) = 0$$

and the associated ordinary differential equation

$$\begin{cases} \dot{X}(t,x) = b(t, X(t,x)) \\ X(0,x) = x. \end{cases}$$

What one proves is that, in a suitable sense (see [56, 4, 5] for a precise statement), existence and uniqueness for the ordinary differential equation hold for almost every initial condition if, and only if, the partial differential equation is well-posed in L^∞. It was pointed out in [4] that this theory has a probabilistic flavour, and therefore it is very natural to look for a more general theory concerning *stochastic* differential equations whose limit, as the diffusion coefficient tends to 0, should be the DiPerna-Lions theory.

In Section 5.2 we obtain this type of extension [77]: first we study the links between the *Fokker-Planck equation*

$$\partial_t \mu_t + \sum_i \partial_i (b_i \mu_t) - \frac{1}{2} \sum_{ij} \partial_{ij} (a_{ij} \mu_t) = 0 \quad \text{in } [0,T] \times \mathbb{R}^d, \tag{0.0.4}$$

and the associated stochastic differential equation

$$\begin{cases} dX = b(t,X)\, dt + \sigma(t,X)\, dB(t) \\ X(0) = x, \end{cases} \tag{0.0.5}$$

where $a_{ij} := \sum_k \sigma_{ik}\sigma_{jk}$. The stochastic differential equation is considered in a weak sense (that of martingale solutions), and we show that existence and uniqueness for the stochastic differential equation hold for almost every initial condition if and only if the partial differential equation is well-posed in L^∞ (again uniqueness for the stochastic differential equation holds in a more complicated sense). Moreover, a study of the Fokker-Planck equations from a purely partial differential equation point of view (see Section 5.4) allows to apply the above theory to some important specific cases.

As we already said, all these chapters stem from a series of papers published during the PhD studies, except some parts of Chapter 1. Indeed, a preliminary version of the work with Albert Fathi [66] with weaker results was already present in the undergraduate thesis [73]. Also the work with Cédric Villani [78] and the paper [75] were present in [73] but we decided to include them in the chapter for a more complete and organic exposition. We have chosen to leave out, since they are not directly linked with the thesis' subject, some other papers written during the PhD studies (see [76] and [1]). A paper somehow related to optimal transportation, written before the beginning of the PhD studies, is instead [74].

Regarding notation, we tried to make it as much unified as possible. Nevertheless, the main specific notation will be introduced chapter by chapter.

Chapter 1

The optimal transportation problem

1.1 Introduction

[1] The optimal transportation problem we consider in this chapter is the following: given two probability measures μ and ν, defined on the measurable spaces X and Y, find a measurable map $T : X \to Y$ with

$$T_\sharp \mu = \nu, \tag{1.1.1}$$

and in such a way that T minimize the transportation cost, that is

$$\int_X c(x, T(x))\, d\mu(x) = \min_{S_\sharp \mu = \nu} \left\{ \int_X c(x, S(x))\, d\mu(x) \right\}.$$

Here $c : X \times Y \to \mathbb{R}$ is some given cost function, and the minimum is taken over all measurable maps $S : X \to Y$ with $S_\sharp \mu = \nu$. When condition (1.1.1) is satisfied, we say that T is a *transport map*, and if T minimize also the cost we call it an *optimal transport map*.

Even in Euclidean spaces, and the cost c equal to the Euclidean distance or its square, the problem of the existence of an optimal transport map is far from being trivial. Due to the strict convexity of the square of the Euclidean distance, the case $c(x, y) = |x - y|^2$ is simpler to deal with than the case $c(x, y) = |x - y|$. The reader should consult the books and surveys given above to have a better view of the history of the subject, in particular Villani's second book on the subject [133]. However for the case where the cost is a distance, one should cite at least the work of Sudakov [129],

[1]This chapter is based on joint works with Albert Fathi [66], Cédric Villani [78], and on the work in [75].

Evans-Gangbo [60], Feldman-McCann [71], Caffarelli-Feldman-McCann [39], Trudinger-Wang [130], Ambrosio-Pratelli [13], and Bernard-Buffoni [23]. For the case where the cost is the square of the Euclidean or of a Riemannian distance, one should cite at least the work of Knott-Smith [87], Brenier [30], Rachev-Rüschendorf [117], Gangbo-McCann [80], McCann [109], and Bernard-Buffoni [22].

Our work is related to the case where the cost behaves like a square of a Riemannian distance. It is strongly inspired by the work of Bernard-Buffoni [22]. In fact, we prove the non-compact version of this last work adapting some techniques that were first used in the Euclidean case in [11] by Ambrosio, Gigli, and Savaré. We show that the Monge transport problem can be solved for the square distance on any complete Riemannian manifold *without any assumption on the compactness or curvature*, with the usual restriction on the measures. Most of the arguments in this chapter are well known to specialists, at least in the compact case, but they have not been put together before and adapted to the case we treat. Of course, there is a strong intersection with some of the results that appear in [133]. For the case where the cost behaves like the distance of a complete non-compact Riemannian manifold, see [74].

We will prove a generalization of the following theorem (see Theorems 1.4.2 and 1.4.3):

Theorem 1.1.1. *Suppose that M is a connected complete Riemannian manifold, whose Riemannian distance is denoted by d. Suppose that $r > 1$. If μ and ν are probability (Borel) measures on M, with μ absolutely continuous with respect to Lebesgue measure, and*

$$\int_M d^r(x, x_0)\, d\mu(x) < \infty \quad \text{and} \quad \int_M d^r(x, x_0)\, d\nu(x) < \infty$$

for some given $x_0 \in M$, then we can find a transport map $T : M \to M$, with $T_\sharp\mu = \nu$, which is optimal for the cost d^r on $M \times M$. Moreover, the map T is uniquely determined μ-almost everywhere.

We recall that a measure on a smooth manifold is absolutely continuous with respect to the Lebesgue measure if, when one looks at it in charts, it is absolutely continuous with respect to Lebesgue measure. Again we note that there is no restriction on the curvature of M in the theorem above.

The chapter is structured as follows: in Section 1.2 we recall some known results on the general theory of the optimal transport problem, and we introduce some useful definitions. Then, in Section 1.3 we will give very general results for the existence and the uniqueness of optimal transport maps (Theorems 1.3.1 and 1.3.2, and Complement 1.3.4). In Section 1.4 the above results are applied in the case of costs functions coming from (weak) Tonelli Lagrangians (Theorems 1.4.2 and 1.4.3). In Section 1.5, we study the so called "dispacement interpolation", showing a countably Lipschitz regularity for the transport map starting from an intermidiate time (Theorem 1.5.2). All the tecnical results about semi-concave functions and Tonelli Lagrangians used in our proofs are collected in the appendix at the end of the thesis.

1.2 Background and some definitions

The weak formulation of the transport problem proposed by Kantorovich in [85], [86] is the following: one looks for *plans* instead of transport maps, that is probability measures γ in $X \times Y$ whose marginals are μ and ν, and one minimizes

$$C(\mu,\nu) = \min_{\gamma \in \Pi(\mu,\nu)} \left\{ \int_{M\times M} c(x,y)\, d\gamma(x,y) \right\}, \tag{1.2.1}$$

(here $\Pi(\mu,\nu)$ denotes the set of plans). If γ is a minimizer for the Kantorovich formulation, we say that it is an *optimal plan*. Using weak topologies, it is simple to prove existence of optimal plans whenever X and Y are Polish spaces and c is lower semicontinuous (see [118], [132, Proposition 2.1] or [133]).

It is well-known that a linear minimization problem with convex constraints, like (1.2.1), admits a dual formulation. Before stating the duality formula, we make some definitions similar to that of the weak KAM theory (see [65]):

Definition 1.2.1 (c-subsolution). We say that a pair of functions $\varphi : X \to \mathbb{R}\cup\{+\infty\}$, $\psi : Y \to \mathbb{R}\cup\{-\infty\}$ is a *c-subsolution* if

$$\forall (x,y) \in X \times Y, \quad \psi(y) - \varphi(x) \le c(x,y).$$

Observe that when c is measurable and bounded below, and (φ,ψ) is a c-subsolution with $\varphi \in L^1(\mu), \psi \in L^1(\nu)$, then

$$\begin{aligned}\forall \gamma \in \Pi(\mu,\nu), \quad \int_Y \psi\, d\nu - \int_X \varphi\, d\mu &= \int_{X\times Y} (\psi(y) - \varphi(x))\, d\gamma(x,y) \\ &\le \int_{X\times Y} c(x,y)\, d\gamma(x,y).\end{aligned}$$

If moreover $\int_{X\times Y} c(x,y)\, d\gamma < +\infty$, and

$$\int_{X\times Y} (\psi(y) - \varphi(x))\, d\gamma(x,y) = \int_{X\times Y} c(x,y)\, d\gamma(x,y),$$

then one would obtain the following equality:

$$\psi(y) - \varphi(x) = c(x,y) \quad \text{for } \gamma\text{-a.e. } (x,y)$$

(without any measurability or integrability assumptions on (φ,ψ), this is just a formal computation).

Definition 1.2.2 (Calibration). Given an optimal plan γ, we say that a c-subsolution (φ, ψ) is (c, γ)-*calibrated* if φ and ψ are Borel measurable, and

$$\psi(y) - \varphi(x) = c(x, y) \quad \text{for } \gamma\text{-a.e. } (x, y).$$

Theorem 1.2.3 (Duality formula). *Let X and Y be Polish spaces equipped with probability measures μ and ν respectively, $c : X \times Y \to \mathbb{R}$ a lower semicontinuous cost function bounded from below such that the infimum in the Kantorovitch problem (1.2.1) is finite. Then a transport plan $\gamma \in \Pi(\mu, \nu)$ is optimal if and only if there exists a (c, γ)-calibrated subsolution (φ, ψ).*

For a proof of this theorem see [120] and [133, Theorem 5.9 (ii)].

Here we study Monge's problem on manifolds for a large class of cost functions induced by Lagrangians like in [22], where the authors consider the case of compact manifolds. We generalize their result to arbitrary non-compact manifolds.

Following the general scheme of proof, we will first prove a result on more general costs, see Theorem 1.3.2. In this general result, the fact that the target space for the Monge transport is a manifold is not necessary. So we will assume that only the source space (for the Monge transport map) is a manifold.

Let M be an n-dimensional manifold (Hausdorff and with a countable basis), N a Polish space, $c : M \times N \to \mathbb{R}$ a cost function, μ and ν two probability measures on M and N respectively. We want to prove existence and uniqueness of an optimal transport map $T : M \to N$, under some reasonable hypotheses on c and μ.

One of the conditions on the cost c is given in the following definition:

Definition 1.2.4 (Twist Condition). For a given cost function $c(x, y)$, we define the *skew left Legendre transform* as the partial map

$$\Lambda_c^l : M \times N \to T^*M,$$

$$\Lambda_c^l(x, y) = (x, \frac{\partial c}{\partial x}(x, y)),$$

whose domain of definition is

$$\mathcal{D}(\Lambda_c^l) = \left\{(x, y) \in M \times N \mid \frac{\partial c}{\partial x}(x, y) \text{ exists}\right\}.$$

Moreover, we say that c satisfies the *left twist condition* if Λ_c^l is injective on $\mathcal{D}(\Lambda_c^l)$.

One can define similarly the *skew right Legendre transform* $\Lambda_c^r : M \times N \to T^*N$ by $\Lambda_c^r(x, y) = (y, \frac{\partial c}{\partial y}(x, y)),$. The domain of definition of Λ_c^r is $\mathcal{D}(\Lambda_c^r) = \{(x, y) \in M \times N \mid \frac{\partial c}{\partial x}(x, y) \text{ exists}\}$. We say that c satisfies the *right twist condition* if Λ_c^r is injective on $\mathcal{D}(\Lambda_c^r)$.

The usefulness of these definitions will be clear in the Section 1.4, in which we will treat the case where $M = N$ and the cost is induced by a Lagrangian. This condition has appeared already in the subject. It has been known (explicitly or not) by several people, among them Gangbo (oral communication) and Villani (see [132, page 90]). It is used in [22], since it is always satisfied for a cost coming from a Lagrangian, as will see below. We borrow the terminology "twist condition" from the theory of Dynamical Systems: if $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (x, y) \mapsto h(x, y)$ is C^2, one says that h satisfies the twist condition if there exists a constant $\alpha > 0$ such that $\dfrac{\partial^2 h}{\partial x \partial y} \geq \alpha$ everywhere. In that case both maps $\Lambda_h^l : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}, (x, y) \mapsto (x, \partial h/\partial x(x, y))$ and $\Lambda_h^r : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}, (x, y) \mapsto (y, \partial h/\partial y(x, y))$ are C^1 diffeomorphisms. The twist map $f : \mathbb{R}\times\mathbb{R} \to \mathbb{R}\times\mathbb{R}$ associated to h is determined by $f(x_1, v_1) = (x_2, v_2)$, where $v_1 = -\partial h/\partial x(x_1, x_2), v_2 = \partial h/\partial y(x_1, x_2)$, which means $f(x_1, v_1) = \Lambda_h^r \circ [\Lambda_h^l]^{-1}(x_1, -v_1)$, see [102] or [79].

We now recall some useful measure-theoretical facts that we will need in the sequel.

Lemma 1.2.5. *Let M be an n-dimensional manifold, N be a Polish space, and let $c : M \times N \to \mathbb{R}$ be a measurable function such that $x \mapsto c(x, y)$ is continuous for any $y \in N$. Then the set*

$$\{(x, y) \mid \frac{\partial c}{\partial x}(x, y) \text{ exists}\} \quad \text{is Borel measurable.}$$

Moreover $(x, y) \mapsto \frac{\partial c}{\partial x}(x, y)$ is a Borel function on that set.

Proof. This a standard result in measure theory, we give here just a sketch of the proof.

By the locality of the statement, using charts we can assume $M = \mathbb{R}^n$. Let $T_k : \mathbb{R}^n \to \mathbb{R}^n$ be a dense countable family of linear maps. For any $j, k \in \mathbb{N}$, we consider the Borel function

$$\begin{aligned} L_{j,k}(x, y) :&= \sup_{|h| \in (0, \frac{1}{j})} \frac{|c(x + h, y) - c(x, y) - T_k(h)|}{|h|} \\ &= \sup_{|h| \in (0, \frac{1}{j}), h \in \mathbb{Q}^n} \frac{|c(x + h, y) - c(x, y) - T_k(h)|}{|h|}, \end{aligned}$$

where in the second equality we used the continuity of $x \mapsto c(x, y)$. Then it is not difficult to show that the set of point where $\frac{\partial c}{\partial x}(x, y)$ exists can be written as

$$\{(x, y) \mid \inf_j \inf_k L_{j,k}(x, y) = 0\},$$

which is clearly a Borel set.

To show that $x \mapsto \frac{\partial c}{\partial x}(x,y)$ is Borel, it suffices to note that the partial derivatives

$$\frac{\partial c}{\partial x_i}(x,y) = \lim_{\ell\to\infty} \frac{c(x_1,\dots,x_i+\frac{1}{\ell},\dots,x_n,y)-\varphi_n(x_1,\dots,x_i,\dots,x_n,y)}{1/\ell}$$

are countable limits of continuous functions, and hence are Borel measurable. □

Therefore, by the above lemma, $\mathcal{D}(\Lambda_c^l)$ is a Borel set. If we moreover assume that c satisfies the left twist condition (that is, Λ_c^l is injective on $\mathcal{D}(\Lambda_c^l)$), then one can define

$$(\Lambda_c^l)^{-1} : T^*M \supset \Lambda_c^l(\mathcal{D}(\Lambda_c^l)) \to \mathcal{D}(\Lambda_c^l) \subset M \times N.$$

Then, by the injectivity assumption, one has that $\Lambda_c^l(\mathcal{D}(\Lambda_c^l))$ is still a Borel set, and $(\Lambda_c^l)^{-1}$ is a Borel map (see [51, Proposition 8.3.5 and Theorem 8.3.7], [70]). We can so extend $(\Lambda_c^l)^{-1}$ as a Borel map on the whole T^*M as

$$\Lambda_c^{l,inv}(x,p) = \begin{cases} (\Lambda_c^l)^{-1}(x,p) & \text{if } p \in T_x^*M \cap \Lambda_c^l(\mathcal{D}(\Lambda_c^l)), \\ (x,\bar{y}) & \text{if } p \in T_x^*M \setminus \Lambda_c^l(\mathcal{D}(\Lambda_c^l)), \end{cases}$$

where $\bar{y}$ is an arbitrary point, but fixed point, in N.

1.3 The main result

In order to have general results of existence and uniqueness of transport maps which are sufficiently flexible to be used also in other situations, and to well show where measure-theoretic problems enter in the proof of the existence of the transport map, we will first give a general result where no measures are present (see in the appendix 6.1.3 for the definition of locally semi-concave function and 6.1.7 for the definition of countably $(n-1)$-Lipschitz set).

Theorem 1.3.1. *Let M be a smooth (second countable) manifold, and let N be a Polish space. Assume that the cost $c : M \times N \to \mathbb{R}$ is Borel measurable, bounded from below, and satisfies the following conditions:*

(i) *the family of maps $x \mapsto c(x,y) = c_y(x)$ is locally semi-concave in x locally uniformly in y,*

(ii) *the cost c satisfies the left twist condition.*

Let (φ,ψ) be a c-subsolution, and consider the set $G_{(\varphi,\psi)} \subset M \times N$ given by

$$G_{(\varphi,\psi)} = \{(x,y) \in M \times N \mid \psi(y) - \varphi(x) = c(x,y)\}.$$

We can find a Borel countably $(n-1)$-Lipschitz set $E \subset M$ and a Borel measurable map $T : M \to N$ such that

$$G_{(\varphi,\psi)} \subset \mathrm{Graph}(T) \cup \pi_M^{-1}(E),$$

where $\pi_M : M \times N \to M$ is the canonical projection, and $\mathrm{Graph}(T) = \{(x, T(x)) \mid x \in M\}$ is the graph of T.

In other words, if we define $P = \pi_M\big(G_{(\varphi,\psi)}\big) \subset M$ the part of $G_{(\varphi,\psi)}$ which is above $P \setminus E$ is contained a Borel graph.

More precisely, we will prove that there exist an increasing sequence of locally semi-convex functions $\varphi_n : M \to \mathbb{R}$, with $\varphi \geq \varphi_{n+1} \geq \varphi_n$ on M, and an increasing sequence of Borel subsets C_n such that

- *For $x \in C_n$, the derivative $d_x\varphi_n$ exists, $\varphi_{n+1}(x) = \varphi_n(x)$, and $d_x\varphi_{n+1} = d_x\varphi_n$.*
- *If we set $C = \cup_n C_n$, there exists a Borel countably $(n-1)$-Lipschitz set $E \subset M$ such that $P \setminus E \subset C$.*

Moreover, the Borel map $T : M \to N$ is such that

- *For every $x \in C_n$, we have*

  $$(x, T(x)) = \Lambda_c^{l,inv}(x, -d_x\varphi_n),$$

  where $\Lambda_c^{l,inv}$ is the extension of the inverse of Λ_c^l defined at the end of Section 1.2.
- *If $x \in P \cap C_n \setminus E$, then the partial derivative $\dfrac{\partial c}{\partial x}(x, T(x))$ exists (i.e. $(x, T(x)) \in \mathcal{D}(\Lambda_c^l)$), and*

  $$\frac{\partial c}{\partial x}(x, T(x)) = -d_x\varphi_n.$$

In particular, if $x \in P \cap C_n \setminus E$, we have

$$(x, T(x)) \in \mathcal{D}(\Lambda_c^l) \quad \text{and} \quad \Lambda_c^l(x, T(x)) = (x - d_x\varphi_n).$$

Therefore, thanks to the twist condition, the map T is uniquely defined on $P \setminus E \subset C$.

The existence and uniqueness of a transport map is then a simple consequence of the above theorem.

Theorem 1.3.2. *Let M be a smooth (second countable) manifold, let N be a Polish space, and consider μ and ν (Borel) probability measures on M and N respectively. Assume that the cost $c : M \times N \to \mathbb{R}$ is lower semicontinuous and bounded from below. Assume moreover that the following conditions hold:*

(i) *the family of maps* $x \mapsto c(x,y) = c_y(x)$ *is locally semi-concave in* x *locally uniformly in* y*,*

(ii) *the cost* c *satisfies the left twist condition,*

(iii) *the measure* μ *gives zero mass to countably* $(n-1)$*-Lipschitz sets,*

(iv) *the infimum in the Kantorovitch problem (1.2.1) is finite.*

Then there exists a Borel map $T : M \to N$*, which is an optimal transport map from* μ *to* ν *for the cost* c*. Morover, the map* T *is unique* μ*-a.e., and any plan* $\gamma_c \in \Pi(\mu, \nu)$ *optimal for the cost* c *is concentrated on the graph of* T*.*

More precisely, if (φ, ψ) *is a* (c, γ_c)*-calibrating pair, with the notation of Theorem 1.3.1, there exists an increasing sequence of Borel subsets* B_n*, with* $\mu(\cup_n B_n) = 1$*, such that the map* T *is uniquely defined on* $B = \cup_n B_n$ *via*

$$\frac{\partial c}{\partial x}(x, T(x)) = -d_x \varphi_n \quad \text{on } B_n,$$

and any optimal plan $\gamma \in \Pi(\mu, \nu)$ *is concentrated on the graph of that map* T*.*

We remark that condition (iv) is trivially satisfied if

$$\int_{M \times N} c(x,y)\, d\mu(x)\, d\nu(y) < +\infty.$$

However we needed to stated the above theorem in this more general form in order to apply it in Section 1.5 (see Remark 1.5.3).

Proof of Theorem 1.3.2. Let $\gamma_c \in \Pi(\mu, \nu)$ be an optimal plan. By Theorem 1.2.3 there exists a (c, γ)-calibrated pair (φ, ψ). Consider the set

$$G = G_{(\varphi,\psi)} = \{(x,y) \in M \times N \mid \psi(y) - \varphi(x) = c(x,y)\}.$$

Since both M and N are Polish and both maps φ and ψ are Borel, the subset G is a Borel subset of $M \times N$. Observe that, by the definition of (c, γ_c)-calibrated pair, we have $\gamma_c(G) = 1$.

By Theorem 1.3.1 there exists a Borel countably $(n-1)$-Lipschitz set E such that $G \setminus (\pi_M)^{-1}(E)$ is contained in the graph of a Borel map T. This implies that

$$B = \pi_M \left(G \setminus (\pi_M)^{-1}(E) \right) = \pi_M(G) \setminus E \subset M$$

is a Borel set, since it coincides with $(\mathrm{Id}_M \tilde{\times} T)^{-1} \left(G \setminus (\pi_M)^{-1}(E) \right)$ and the map $x \mapsto \mathrm{Id}_M \tilde{\times} T(x) = (x, T(x))$ is Borel measurable.

Thus, recalling that the first marginal of γ_c is μ, by assumption (iii) we get $\gamma_c((\pi_M)^{-1}(E)) = \mu(E) = 0$. Therefore $\gamma_c(G \setminus (\pi_M)^{-1}(E)) = 1$, so that γ_c is concentrated on the graph of T, which gives the existence of an optimal transport map. Note now that $\mu(B) = \gamma_c(\pi^{-1}(B)) \geq \gamma_c(G \setminus (\pi_M)^{-1}(E)) = 1$. Therefore $\mu(B) = 1$. Since $B = P \setminus E$, where $P = \pi_M(G)$, using the Borel set C_n provided by Theorem 1.3.1, it follows that $B_n = P \cap C_n \setminus E = D \cap C_n$ is a Borel set with $B = \cup_n B_n$. The end of Theorem 1.3.1 shows that T is indeed uniquely defined on B as said in the statement.

Let us now prove the uniqueness of the transport map μ-a.e.. If S is another optimal transport map, consider the measures $\gamma_T = (\mathrm{Id}_M \times T)_\# \mu$ and $\gamma_S = (\mathrm{Id}_M \times S)_\# \mu$. The measure $\bar{\gamma} = \frac{1}{2}(\gamma_T + \gamma_S) \in \Pi(\mu, \nu)$ is still an optimal plan, and therefore must be concentrated on a graph. This implies for instance that $S = T$ μ-a.e., and thus T is the unique optimal transport map. Finally, since any optimal $\gamma \in \Pi(\mu, \nu)$ is concentrated on a graph, we also deduce that any optimal plan is concentrated on the graph of T. $\square$

Proof of Theorem 1.3.1. By definition of c-subsolution, we have $\varphi > -\infty$ everywhere on M, and $\psi < +\infty$ everywhere on N. Therefore, if we define $W_n = \{\psi \leq n\}$, we have $W_n \subset W_{n+1}$, and $\cup_n W_n = N$. Since, by hypothesis (i), $c(x, y) = c_y(x)$ is locally semi-concave in x locally uniformly in y, for each $y \in N$ there exist a neighborhood V_y of y such that the family of functions $(c(\cdot, z))_{z \in V_y}$ is locally uniformly semi-concave. Since N is separable, there exists a countable family of points $(y_k)_{k \in \mathbb{N}}$ such that $\cup_k V_{y_k} = N$. We now consider the sequence of subsets $(V_n)_{n \in \mathbb{N}} \subset N$ defined as

$$V_n = W_n \cap \left(\cup_{1 \leq k \leq n} V_{y_k}\right).$$

We have $V_n \subset V_{n+1}$. Define $\varphi_n : M \to N$ by

$$\varphi_n(x) = \sup_{y \in V_n} \psi(y) - c(x, y) = \max_{1 \leq k \leq n} \Big(\sup_{y \in W_n \cap V_{y_k}} \psi(y) - c(x, y) \Big).$$

Since $\psi \leq n$ on K_n, and $-c$ is bounded from above, we see that φ_n is bounded from above. Therefore, by hypothesis (i), the family of functions $(\psi(y) - c(\cdot, y))_{y \in W_n \cap V_{y_k}}$ is locally uniformly semi-convex and bounded from above. Thus, by Theorem 6.1.4 and Proposition 6.1.11 of the Appendix, the function φ_n is locally semi-convex. Since $\psi(y) - \varphi(x) \leq c(x, y)$ with equality on $G_{(\varphi, \psi)}$, and $V_n \subset V_{n+1}$, we clearly have

$$\varphi_n \leq \varphi_{n+1} \leq \varphi \quad \text{everywhere on } M.$$

A key observation is now the following:

$$\varphi|_{P_n} = \varphi_n|_{P_n},$$

where $P_n = \pi_M(G_{(\varphi,\psi)} \cap (M \times V_n))$. In fact, if $x \in P_n$, by the definition of P_n we know that there exists a point $y_x \in V_n$ such that $(x, y_x) \in G_{(\varphi,\psi)}$. By the definition of $G_{(\varphi,\psi)}$, this implies

$$\varphi(x) = \psi(y_x) - c(x, y_x) \leq \varphi_n(x) \leq \varphi(x).$$

Since φ_n is locally semi-convex, by Theorem 6.1.8 of the Appendix applied to $-\varphi_n$, it is differentiable on a Borel subset F_n such that its complement F_n^c is a Borel countably $(n-1)$-Lipschitz set. Let us then define $F = \cap_n F_n$. The complement $E = F^c = \cup_n F_n^c$ is also a Borel countably $(n-1)$-Lipschitz set. We now define the Borel set

$$C_n = F \cap \{x \in M \mid \varphi_k(x) = \varphi_n(x) \ \ \forall k \geq n\}.$$

We observe that $C_n \supset P_n \cap F$.

We now prove that $G_{(\varphi,\psi)} \cap ((P_n \cap F) \times V_n)$ is contained in a graph.

To prove this assertion, fix $x \in P_n \cap F$. By the definition of P_n, and what we said above, there exists $y_x \in V_n$ such that

$$\varphi(x) = \varphi_n(x) = \psi(y_x) - c(x, y_x).$$

Since $x \in F$, the map $z \mapsto \varphi_n(z) - \psi(y_x)$ is differentiable at x. Moreover, by condition (i), the map $z \mapsto -c(z, y_x) = -c_{y_x}(z)$ is locally semi-convex and, by the definition of φ_n, for every $z \in M$, we have $\varphi_n(z) - \psi(y_x) \geq -c(z, y_x)$, with equality at $z = x$. These facts taken together imply that $\frac{\partial c}{\partial x}(x, y_x)$ exists and is equal to $-d_x\varphi_n$. In fact, working in a chart around x, since $c_{y_x} = c(\cdot, y_x)$ is locally semi-concave, by the definition 6.1.3 of a locally semi-concave function, there exists linear map l_x such that

$$c(z, y_x) \leq c(x, y_x) + l_x(z - x) + o(|z - x|),$$

for z in a neighborhood of x. Using also that φ_n is differentiable at x, we get

$$\begin{aligned}\varphi_n(x) - \psi(y_x) + d_x\varphi_n(z - x) + o(|z - x|) &= \varphi_n(z) - \psi(y_x) \\ &\geq -c(z, y_x) \geq -c(x, y_x) - l_c(z - x) + o(|z - x|) \\ &= \varphi_n(x) - \psi(y_x) - l_c(z - x) + o(|z - x|).\end{aligned}$$

This implies that $l_c = -d_x\varphi_n$, and that c_{y_x} is differentiable at x with differential at x equal to $-d_x\varphi_n$. Setting now $G_x = \{y \in N \mid \varphi(x) - \psi(y) = c(x, y)\}$, we have just shown that $\{x\} \times (G_x \cap V_n) \subset \mathcal{D}(\Lambda_c^l)$ for each $x \in C_n$, and also $\frac{\partial c}{\partial x}(x, y) = -d_x\varphi_n$, for every $y \in G_x \cap V_n$. Recalling now that that, by hypothesis (ii), the cost c satisfies the left twist condition, we obtain that $G_x \cap V_n$ is reduced to a single element which is uniquely characterized by the equality

$$\frac{\partial c}{\partial x}(x, y_x) = -d_x\varphi_n.$$

So we have proved that $G \cap (M \times V_n)$ is the graph over $P_n \cap F$ of the map T defined uniquely, thanks to the left twist condition, by

$$\frac{\partial c}{\partial x}(x, T(x)) = -d_x\varphi_n$$

(observe that, since $\varphi_n \leq \varphi_k$ for $k \geq n$ with equality on P_n, we have $d_x\varphi_n|_{P_n} = d_x\varphi_k|_{P_n}$ for $k \geq n$). Since $P_{n+1} \supset P_n$, and $V_n \subset V_{n+1} \nearrow N$, we can conclude that $G_{(\varphi,\psi)}$ is a graph over $\cup_n P_n \cap F = P \cap F$ (where $P = \pi_M(G_{(\varphi,\psi)}) = \cup_n P_n$).

Observe that, at the moment, we do not know that T is a Borel map, since P_n is not a priori Borel. Note first that by definition of $B_n \subset B_{n+1}$, we $\varphi_n = \varphi_{n+1}$ on B_n, and they are both differentiable at every point of B_n. Since $\varphi_n \leq \varphi_{n+1}$ everywhere, by the same argument as above we get $d_x\varphi_n = d_x\varphi_{n+1}$ for $x \in B_n$. Thus, setting $B = \cup_n B_n$, we can extend T to M by

$$T(x) = \begin{cases} \pi_N \Lambda_c^{l,inv}(x, -d_x\varphi_n) & \text{on } B_n, \\ \bar{y} & \text{on } M \setminus B, \end{cases}$$

where $\pi_N : M \times N \to N$ is the canonical projection, $\Lambda_c^{l,inv}$ is the Borel extension of $(\Lambda_c^l)^{-1}$ defined after Lemma 1.2.5, and $\bar{y}$ is an arbitrary but fixed point in N. Obviously, the map T thus defined is Borel measurable and extends the map T already defined on $P \setminus E$. □

In the case where μ is absolutely continuous with respect to Lebesgue measure we can give a complement to our main theorem. In order to state it, we need the following definition, see [11, Definition 5.5.1, page 129]:

Definition 1.3.3 (Approximate differential). We say that $f : M \to \mathbb{R}$ has an *approximate differential* at $x \in M$ if there exists a function $h : M \to \mathbb{R}$ differentiable at x such that the set $\{f = h\}$ has density 1 at x with respect to the Lebesgue measure (this just means that the density is 1 in charts). In this case, the approximate value of f at x is defined as $\tilde{f}(x) = h(x)$, and the approximate differential of f at x is defined as $\tilde{d}_x f = d_x h$. It is not difficult to show that this definition makes sense. In fact, both $h(x)$, and $d_x h$ do not depend on the choice of h, provided x is a density point of the set $\{f = h\}$.

Another characterization of the approximate value $\tilde{f}(x)$, and of the approximate differential $\tilde{d}_x f$ is given, in charts, saying that the sets

$$\left\{ y \mid \frac{|f(y) - \tilde{f}(x) - \tilde{d}_x f(y - x)|}{|y - x|} > \varepsilon \right\}$$

have density 0 at x for each $\varepsilon > 0$ with respect to Lebesgue measure. This last definition is the one systematically used in [70]. On the other hand, for our purpose, Definition 1.3.3 is more convenient.

The set points $x \in M$ where the approximate derivative $\tilde{d}_x f$ exists is measurable; moreover, the map $x \mapsto \tilde{d}_x f$ is also measurable, see [70, Theorem 3.1.4, page 214].

Complement 1.3.4. *Under the hypothesis of Theorem 1.3.2, if we assume that μ is absolutely continuous with respect to Lebesgue measure (this is stronger than condition (iii) of Theorem 1.3.2), then for any calibrated pair (φ, ψ), the function φ is approximatively differentiable μ-a.e., and the optimal transport map T is uniquely determined μ-a.e., thanks to the twist condition, by*

$$\frac{\partial c}{\partial x}(x, T(x)) = -\tilde{d}_x \varphi,$$

where $\tilde{d}_x \varphi$ is the approximate differential of φ at x. Moreover, there exists a Borel subset $A \subset M$ of full μ measure such that $\tilde{d}_x \varphi$ exists on A, the map $x \mapsto \tilde{d}_x \varphi$ is Borel measurable on A, and $\frac{\partial c}{\partial x}(x, T(x))$ exists for $x \in A$ (i.e. $(x, T(x)) \in \mathcal{D}(\Lambda_c^l)$).

Proof. We will use the notation and the proof of Theorems 1.3.1 and 1.3.2. We denote by $\tilde{A}_n \subset B_n$ the set of $x \in B_n$ which are density points for B_n with respect to some measure λ whose measure class in charts is that of Lebesgue (for example one can take λ as the Riemannian measure associated to a Riemannian metric). By Lebesgue's Density Theorem $\lambda(B_n \setminus \tilde{A}_n) = 0$. Since μ is absolutely continuous with respect to Lebesgue measure, we have $\mu(\tilde{A}_n) = \mu(B_n)$, and therefore $\tilde{A} = \cup_n \tilde{A}_n$ is of full μ-measure, since $\mu(B_n) \nearrow \mu(B) = 1$. Moreover, since $\{\varphi = \varphi_n\}$ on B_n, and φ_n is differentiable at each point of B_n, the function φ is approximatively differentiable at each point of $\tilde{A}_n$ with $\tilde{d}_x \varphi = d_x \varphi_n$.

The last part of this complement on measurability follows of course from [70, Theorem 3.1.4, page 214]. But in this case, we can give a direct simple proof. We choose $A_n \subset \tilde{A}_n$ Borel measurable with $\mu(\tilde{A}_n \setminus A_n) = 0$. We set $A = \cup_n A_n$. The set A is of full μ measure. Moreover, for every $x \in A_n$, the approximate differential $\tilde{d}_x \varphi$ exists and is equal to $d_x \varphi_n$. Thus it suffices to show that the map $x \mapsto d_x \varphi_n$ is Borel measurable, and this follows as in Lemma 1.2.5. □

1.4 Costs obtained from Lagrangians

Now that we have proved Theorem 1.3.2, we want to observe that the hypotheses are satisfied by a large class of cost functions.

We will consider first the case of a Tonelli Lagrangian L on a connected manifold (see Definition 6.2.4 of the Appendix for the definition of a Tonelli Lagrangian). For $t > 0$, the cost $c_{t,L} : M \times M \to \mathbb{R}$ associated to L is given by

$$c_{t,L}(x, y) = \inf_\gamma \int_0^t L(\gamma(s), \dot\gamma(s))\, ds,$$

where the infimum is taken over all the continuous piecewise C^1 curves $\gamma : [0, t] \to M$, with $\gamma(0) = x$, and $\gamma(t) = y$ (see Definition 6.2.18 of the Appendix).

Proposition 1.4.1. *If $L : TM \to \mathbb{R}$ is a Tonelli Lagrangian on the connected manifold M, then, for $t > 0$, the cost $c_{t,L} : M \times M \to \mathbb{R}$ associated to the Lagrangian L is continuous, bounded from below, and satisfies conditions* (i) *and* (ii) *of Theorem 1.3.2.*

Proof. Since L is a Tonelli Lagrangian, observe that L is bounded below by C, where C is the constant given in condition (c) of Definition 6.2.4. Hence the cost $c_{t,L}$ is bounded below by tC. By Theorem 6.2.19 of the Appendix, the cost $c_{t,L}$ is locally semi-concave, and therefore continuous. Moreover, we can now apply Proposition 6.1.17 of the Appendix to conclude that $c_{t,L}$ satisfies condition (i) of Theorem 1.3.2.

The twist condition (ii) of Theorem 1.3.2 for $c_{t,L}$ follows from Lemma 6.2.22 and Proposition 6.2.23. □

For costs coming from Tonelli Lagrangians, we subsume the application of the main Theorem 1.3.2, and its Complement 1.3.4.

Theorem 1.4.2. *Let L be a Tonelli Lagrangian on the connected manifold M. Fix $t > 0$, μ, ν a pair of probability measure on M, with μ giving measure zero to countably $(n-1)$-Lipschitz sets, and assume that the infimum in the Kantorovitch problem (1.2.1) with cost $c_{t,L}$ is finite. Then there exists a uniquely μ-almost everywhere defined transport map $T : M \to M$ from μ to ν which is optimal for the cost $c_{t,L}$. Moreover, any plan $\gamma \in \Pi(\mu, \nu)$, which is optimal for the cost $c_{t,L}$, verifies $\gamma(\mathrm{Graph}(T)) = 1$.*

If μ is absolutely continuous with respect to Lebesgue measure, and (φ, ψ) is a $c_{t,L}$-calibrated subsolution for (μ, ν), then we can find a Borel set B of full μ measure, such that the approximate differential $\tilde d_x\varphi$ of φ at x is defined for $x \in B$, the map $x \mapsto \tilde d_x\varphi$ is Borel measurable on B, and the transport map T is defined on B (hence μ-almost everywhere) by

$$T(x) = \pi^*\phi_t^H(x, \tilde d_x\varphi),$$

where $\pi^* : T^*M \to M$ is the canonical projection, and ϕ_t^H is the Hamiltonian flow of the Hamiltonian H associated to L.

We can also give the following description for T valid on B (hence μ-almost everywhere):

$$T(x) = \pi\phi_t^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi)),$$

where ϕ_t^L is the Euler-Lagrange flow of L, and $x \to \widetilde{\operatorname{grad}}_x^L(\varphi)$ is the measurable vector field on M defined on B by

$$\frac{\partial L}{\partial v}(x, \widetilde{\operatorname{grad}}_x^L(\varphi)) = \tilde{d}_x\varphi.$$

Moreover, for every $x \in B$, there is a unique L-minimizer $\gamma : [0,t] \to M$, with $\gamma(0) = x, \gamma(t) = T(x)$, and this curve γ is given by $\gamma(s) = \pi\phi_s^L(x, \widetilde{\operatorname{grad}}_x^L(\varphi))$, for $0 \leq s \leq t$.

Proof. The first part is a consequence of Proposition 1.4.1 and Theorem 1.3.2. When μ is absolutely continuous with respect to Lebesgue measure, we can apply Complement 1.3.4 to obtain a Borel subset $A \subset M$ of full μ measure such that, for every $x \in A$, we have $(x, T(x)) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$ and

$$\frac{\partial c_{t,L}}{\partial x}(x, T(x)) = \tilde{d}_x\varphi.$$

By Lemma 6.2.22 and Proposition 6.2.23, if $(x, y) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$, then there is a unique L-minimizer $\gamma : [0,t] \to M$, with $\gamma(0) = x, \gamma(t) = y$, and this minimizer is of the form $\gamma(s) = \pi\phi_s^L(x, v)$, where $\pi : TM \to M$ is the canonical projection, and $v \in T_xM$ is uniquely determined by the equation

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, v).$$

Therefore $T(x) = \pi\phi_t^L(x, \widetilde{\operatorname{grad}}_x^L(\varphi))$, where $\widetilde{\operatorname{grad}}_x^L(\varphi)$ is uniquely determined by

$$\frac{\partial L}{\partial v}(x, \widetilde{\operatorname{grad}}_x^L(\varphi)) = -\frac{\partial c_{t,L}}{\partial x}(x, T(x)) = \tilde{d}_x\varphi,$$

which is precisely the second description of T. The first description of T follows from the second one, once we observe that

$$\begin{aligned} \mathscr{L}(x, \widetilde{\operatorname{grad}}_x^L(\varphi))) &= (x, \frac{\partial L}{\partial v}(x, \widetilde{\operatorname{grad}}_x^L(\varphi)) = (x, \tilde{d}_x\varphi) \\ \phi_t^H &= \mathscr{L} \circ \phi_t^L \circ \mathscr{L}^{-1} \\ \pi^* \circ \mathscr{L} &= \pi, \end{aligned}$$

where $\mathscr{L} : TM \to T^*M$ is the global Legendre Transform, see Definition 6.2.8 of the Appendix. □

We now turn to the proof of Theorem 1.1.1, which is not a consequence of Theorem 1.4.2 since the cost d^r with $r > 1$ does not come from a Tonelli Lagrangian for $r \neq 2$.

Theorem 1.4.3. *Suppose that the connected manifold M is endowed with a Riemannian metric g which is complete. Denote by d the Riemannian distance. If $r > 1$, and μ and ν are probability (Borel) measures on M, where μ gives measure zero to countably $(n-1)$-Lipschitz sets, and*

$$\int_M d^r(x, x_0)\, d\mu(x) < \infty \quad \text{and} \quad \int_M d^r(x, x_0)\, d\nu(x) < \infty$$

for some given $x_0 \in M$, then we can find a transport map $T : M \to M$, with $T_\sharp\mu = \nu$, which is optimal for the cost d^r on $M \times M$. Moreover, the map T is uniquely determined μ-almost everywhere.

If μ is absolutely continuous with respect to Lebesgue measure, and (φ, ψ) is a calibrated subsolution for the cost $d^r(x, y)$ and the pair of measures (μ, ν), then the approximate differential $\tilde{d}_x\varphi$ of φ at x is defined μ-almost everywhere, and the transport map T is defined μ-almost everywhere by

$$T(x) = \exp_x\Big(\frac{\tilde{\nabla}_x\varphi}{r^{1/(r-1)}\|\tilde{\nabla}_x\varphi\|_x^{(r-2)/(r-1)}}\Big),$$

where the approximate Riemannian gradient $\tilde{\nabla}\varphi$ of φ is defined by

$$g_x(\tilde{\nabla}_x\varphi, \cdot) = \tilde{d}_x\varphi,$$

and $\exp : TM \to M$ is the exponential map of g on TM, which is globally defined since M is complete.

Proof. We first remark that

$$\begin{aligned} d^r(x, y) &\le [d(x, x_0) + d(x_0, y)]^r \\ &\le [2\max(d(x, x_0), d(x_0, y))]^r \\ &\le 2^r[d(x, x_0)^r + d(y, x_0)^r]. \end{aligned}$$

Therefore

$$\begin{aligned} \int_{M\times M} d^r(x, y)\, d\mu(x)d\nu(y) &\le \int_{M\times M} 2^r[d(x, x_0)^r + d(y, x_0)^r]\, d\mu(x)d\nu(y) \\ &= 2^r\int_M d^r(x, x_0)\, d\mu(x) + 2^r\int_M d^r(y, x_0)\, d\nu(y) < \infty, \end{aligned}$$

and thus the infimum in the Kantorovitch problem (1.2.1) with cost d^r is finite.

By Example 6.2.5, the Lagrangian $L_{r,g}(x, v) = \|v\|_x^r = g_x(v, v)^{r/2}$ is a weak Tonelli Lagrangian. By Proposition 6.2.24, the non-negative and continuous cost $d^r(x, y)$ is precisely the cost $c_{1,L_{r,g}}$. Therefore this cost is locally semi-concave by Theorem 6.2.19. By Proposition 6.1.17, this implies that $d^r(x,y)$ satisfies condition (i) of Theorem 1.3.2. The fact that the cost $d^r(x,y)$ satisfies the left twist condition (ii) of Theorem 1.3.2 follows from Proposition 6.2.24. Therefore there is an optimal transport map T.

If the measure μ is absolutely continuous with respect to Lebesgue measure, and (φ,ψ) is a calibrated subsolution for the cost $d^r(x,y)$ and the pair of measures (μ,ν), then by Complement 1.3.4, for μ-almost every x, we have $(x,T(x)) \in \mathcal{D}(\Lambda^l_{c_{1,L_{r,g}}})$, and

$$\frac{\partial c_{t,L_{r,g}}}{\partial x}(x,T(x)) = -\tilde{d}_x\varphi.$$

Since $(x,T(x))$ is in $\mathcal{D}(\Lambda^l_{c_{1,L_{r,g}}})$, it follows from Proposition 6.2.24 that $T(x) = \pi\phi^g_1(x,v_x)$, where $\pi : TM \to M$ is the canonical projection, the flow ϕ^g_t is the geodesic flow of g on TM, and $v_x \in T_xM$ is determined by

$$\frac{\partial c_{t,L_{r,g}}}{\partial x}(x,T(x)) = -\frac{\partial L_{r,g}}{\partial v}(x,v_x),$$

or, given the equality above, by

$$\frac{\partial L_{r,g}}{\partial v}(x,v_x) = \tilde{d}_x\varphi.$$

Now the vertical derivative of $L_{r,g}$ is computed in Example 6.2.5

$$\frac{\partial L_{r,g}}{\partial v}(x,v) = r\|v\|_x^{r-2} g_x(v,\cdot).$$

Hence $v_x \in T_xM$ is determined by

$$r\|v_x\|_x^{r-2} g_x(v_x,\cdot) = \tilde{d}_x\varphi = g_x(\tilde{\nabla}_x\varphi,\cdot).$$

This gives the equality

$$r\|v_x\|_x^{r-2} v_x = \tilde{\nabla}_x(\varphi),$$

from which we easily get

$$v_x = \frac{\tilde{\nabla}_x\varphi}{r^{1/(r-1)}\|\tilde{\nabla}_x\varphi\|_x^{(r-2)/(r-1)}}.$$

Therefore

$$T(x) = \pi\phi^g_t\Big(x, \frac{\tilde{\nabla}_x\varphi}{r^{1/(r-1)}\|\tilde{\nabla}_x\varphi\|_x^{(r-2)/(r-1)}}\Big).$$

By definition of the exponential map $\exp : TM \to M$, we have $\exp_x(v) = \pi\phi^g_t(x,v)$, and the formula for $T(x)$ follows. $\square$

1.5 The interpolation and its absolute continuity

For a cost $c_{t,L}$ coming from a Tonelli Lagrangian L, Theorem 1.4.2 shows not only that we have an optimal transport map T but also that this map is obtained by following an extremal for time t. We can therefore interpolate the optimal transport by maps T_s where we stop at intermediary times $s \in [0,t]$. We will show in this section that these maps are also optimal transport maps for costs coming from the same Lagrangian. Let us give now precise definitions.

For the sequel of this section, we consider L a Tonelli Lagrangian on the connected manifold M. We fix $t > 0$ and μ_0 and μ_t two probability measures on M, with μ_0 absolutely continuous with respect to Lebesgue measure, and such that

$$\min_{\gamma\in\Pi(\mu_0,\mu_t)} \left\{ \int_{M\times M} c_{t,L}(x,y)\, d\gamma(x,y) \right\} < +\infty.$$

We call T_t the optimal transport map given by Theorem 1.4.2 for $(c_{t,L},\mu_0,\mu_t)$. We denote by (φ,ψ) a fixed $(c_{t,L},\gamma_t)$-calibrated pair. Therefore $\gamma_t = (\mathrm{Id}_M \tilde{\times} T_t)_\# \mu_0$ is the unique optimal plan from μ_0 to μ_t. By Theorem 1.4.2, we can find a Borel subset $B \subset M$ such that:

- the subset B is of full μ_0 measure;
- the approximate ddifferential $\tilde{d}_x\varphi$ exists for every $x \in B$, and is Borel measurable on B;
- the map T_t is defined at every $x \in B$, and we have

$$T_t(x) = \pi\phi_t^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi)),$$

  where ϕ_L^t is the Euler-Lagrange flow of L, $\pi : TM \to M$ is the canonical projection, and the Lagrangian approximate gradient $x \mapsto \widetilde{\mathrm{grad}}_x^L(\varphi))$ is defined by

$$\frac{\partial L}{\partial v}(x, \widetilde{\mathrm{grad}}_x^L(\varphi)) = \tilde{d}_x\varphi;$$

- for every $x \in B$, the partial derivative $\dfrac{\partial c}{\partial x}(x, T_t(x))$ exists, and is uniquely defined by

$$\frac{\partial c}{\partial x}(x, T_t(x)) = -\tilde{d}_x\varphi;$$

- for every $x \in B$ there exists a unique L-minimizer $\gamma_x : [0,t] \to M$ between x and $T_t(x)$. This L-minimizer γ_x is given by

$$\forall s \in [0,t], \quad \gamma_x(s) = \pi\phi_s^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi));$$

- for every $x \in B$, we have

$$\psi(T_t(x)) - \varphi(x) = c_{t,L}(x, T_t(x)).$$

We now make the following important remark, that we will need also in the sequel:

Remark 1.5.1. We observe that, for μ_0-a.e. x, there exists an unique curve from x to $T_t(x)$ that minimizes the action. In fact, since $\frac{\partial c}{\partial x}(x,y)$ exists at $y = T_t(x)$ for μ_0-a.e. x, the twist conditions proved in Section 1.4 tells us that its velocity at time 0 is μ_0-a.e. univocally determined.

For $s \in [0,t]$, we can therefore define μ_0-a.e. an interpolation $T_s : M \to M$ by

$$\forall x \in B, \quad T_s(x) = \gamma_x(s) = \pi\phi_s^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi)).$$

Each map T_s is Borel measurable. In fact, since the global Legendre transform is a homeomorphism and the approximate differential is Borel measurable, the Lagrangian approximate gradient $\widetilde{\mathrm{grad}}^L(\varphi)$ is itself Borel measurable. Moreover the map $\pi\phi_s^L : TM \to M$ is continuous, and thus T_s is Borel measurable. We can therefore define the probability measure $\mu_s = T_{s\#}\mu_0$ on M, i.e. the measure μ_s is the image of μ_0 under the Borel measurable map T_s.

Theorem 1.5.2. *Under the hypothesises above, the maps T_s satisfies the following properties:*

(i) *For every $s \in (0,t)$, the map T_s is the (unique) optimal transport maps for the cost $c_{s,L}$ and the pair of measures (μ_0, μ_s).*

(ii) *For every $s \in (0,t)$, the map $T_s : B \to M$ is injective. Moreover, if we define $\bar{c}_{s,L}(x,y) = c_{s,L}(y,x)$, the inverse map $T_s^{-1} : T_s(B) \to B$ is the (unique) optimal transport map for the cost $\bar{c}_{s,L}$ and the pair of measures (μ_s, μ_0), and it is countably Lipschitz (i.e. there exist a Borel countable partition of M such that T_s^{-1} is Lipschitz on each set).*

(iii) *For every $s \in (0,t)$, the measure $\mu_s = T_{s\#}\mu_0$ is absolutely continuous with respect to Lebesgue measure.*

(iv) *For every $s \in (0,t)$, the composition $\hat{T}_s = T_t T_s^{-1}$ is the (unique) optimal transport map for the cost $c_{t-s,L}$ and the pair of measures (μ_s, μ_t), and it is countably Lipschitz.*

Proof. Fix $s \in (0,t)$. It is not difficult to see, from the definition of $c_{t,L}$, that

$$\forall x, y, z \in M, \quad c_{t,L}(x,y) \leq c_{s,L}(x,y) + c_{t-s,L}(y,z), \tag{1.5.1}$$

and even that

$$\forall x, z \in M, \quad c_{t,l}(x,y) = \inf_{y \in M} c_{s,L}(x,y) + c_{t-s,L}(y,z).$$

If $\gamma : [a,b] \to M$ is an L-minimizer, the restriction $\gamma|_{[c,d]}$ to a subinterval $[c,d] \subset [a,b]$ is also an L-minimizer. In particular, we obtain

$$\forall s \in]a,b[, \quad c_{b-a,L}(\gamma(a),\gamma(b)) = c_{s-a,L}(\gamma(a),\gamma(s)) + c_{b-s,L}(\gamma(s),\gamma(b)).$$

Applying this to the L-minimizer γ_x, we get

$$\forall x \in B, \quad c_{t,L}(x, T_t(x)) = c_{s,L}(x, T_s(x)) + c_{t-s,L}(T_s(x), T_t(x)). \tag{1.5.2}$$

We define for every $s \in (0,t)$, two probability measures $\gamma_s, \hat{\gamma}_s$ on $M \times M$, by

$$\gamma_s = (\mathrm{Id}_M \tilde{\times} T_s)_\# \mu_0 \quad \text{and} \quad \hat{\gamma}_s = (T_s \tilde{\times} T_t)_\# \mu_0,$$

where $\mathrm{Id}_M \tilde{\times} T_s$ and $T_s \tilde{\times} T_t$ are the maps from the subset B of full μ_0 measure to $M \times M$ defined by

$$(\mathrm{Id}_M \tilde{\times} T_s)(x) = (x, T_s(x)),$$
$$(T_s \tilde{\times} T_t)(x) = (T_s(x), T_t(x)).$$

Note that the marginals of γ_s are (μ_0, μ_s), and those of $\hat{\gamma}_s$ are (μ_s, μ_t). We claim that $c_{s,L}(x,y)$ is integrable for γ_s and $\hat{\gamma}_{t-s}$. In fact, we have $C = \inf_{TM} L > -\infty$, hence $c_{r,L} \geq Cr$. Therefore, the equality (1.5.2) gives

$$\forall x \in B, \quad [c_{t,L}(x, T_t(x)) - Ct] = [c_{s,L}(x, T_s(x)) - Cs] + [c_{t-s,L}(T_s(x), T_t(x)) - C(t-s)].$$

Since the functions between brackets are all non-negative, we can integrate this equality with respect to μ_0 to obtain

$$\int_{M \times M} [c_{t,L}(x,y) - Ct] \, d\gamma_t = \int_{M \times M} [c_{s,L}(x,y) - Cs] \, d\gamma_s + \int_{M \times M} [c_{t-s,L}(x,y) - C(t-s)] \, d\hat{\gamma}_s.$$

But all numbers involved in the equality above are non-negative, all measures are probability measures, and the cost $c_{t,L}$ is γ_t integrable since γ_t is an optimal plan for $(c_{t,L}, \mu_0, \mu_t)$, and the optimal cost of $(c_{t,L}, \mu_0, \mu_t)$ is finite. Therefore we obtain that $c_{s,L}$ is γ_s-integrable, and $c_{t-s,L}$ is $\hat{\gamma}_s$-integrable.

Since by definition of a calibrating pair we have $\varphi > -\infty$ and $\psi < +\infty$ everywhere on M, we can find an increasing sequence of compact subsets $K_n \subset M$ with $\cup_n K_n = M$, and we consider $V_n = K_n \cap \{\varphi \geq -n\}$, $V'_n = K_n \cap \{\psi \leq n\}$, so that $\cup_n V_n = \cup_n V'_n = M$.

We define the functions $\varphi_s^n, \psi_s^n : M \to \mathbb{R}$ by

$$\psi_s^n(z) = \inf_{\tilde{z} \in V_n} \varphi(\tilde{z}) + c_{s,L}(\tilde{z}, z),$$
$$\varphi_s^n(z) = \sup_{\tilde{z} \in V'_n} \psi(\tilde{z}) - c_{t-s,L}(z, \tilde{z}),$$

where (φ, ψ) is the fixed $c_{t,L}$-calibrated pair. Note that ψ_s^n is bounded from below by $-n + t \inf_{TM} L > -\infty$. Moreover, the family of functions $(\varphi(\tilde{z}) + c_{s,L}(\tilde{z}, \cdot))_{\tilde{z} \in V'_n}$ is locally uniformly semi-concave with a linear modulus, since this is the case for the family of functions $(c_{s,L}(\tilde{z}, \cdot))_{\tilde{z} \in V'_n}$, by Theorem 6.2.19 and Proposition 6.1.17. It follows from Proposition 6.1.16 that ψ_s^n is semi-concave with a linear modulus. A similar argument proves that $-\varphi_s^n$ is semi-concave with a linear modulus. Note also that, since V_n and V'_n are both increasing sequences, we have $\psi_s^n \geq \psi_s^{n+1}$ and $\varphi_s^{n+1} \leq \varphi_s^n$, for every n. Therefore we can define φ_s (resp. ψ_s) as the pointwise limit of the sequence φ_s^n

Using the fact that (φ, ψ) is a $c_{t,L}$-subsolution, and inequality (1.5.1) above, we obtain

$$\forall x, y, z \in M, \quad \psi(y) - c_{t-s,L}(z, y) \leq \varphi(x) + c_{s,L}(x, z).$$

Therefore we obtain for $x \in V_n, y \in V'_n, z \in M$

$$\psi(y) - c_{t-s,L}(z, y) \leq \varphi_s^n(z) \leq \varphi_s(z) \leq \psi_s(z) \leq \psi_s^n(z) \leq \varphi(x) + c_{s,L}(x, z). \qquad (1.5.3)$$

Inequality (1.5.3) above yields

$$\forall x, y, z \in M, \quad \psi(y) - c_{t-s,L}(z, y) \leq \varphi_s(z) \leq \psi_s(z) \leq \varphi(x) + c_{s,L}(x, z). \qquad (1.5.4)$$

In particular, the pair (φ, ψ_s) is a $c_{s,L}$-subsolution, and the pair (φ_s, ψ) is a $c_{t-s,L}$-subsolution. Moreover, φ, ψ_s, φ_s and ψ are all Borel measurable.

We now define

$$B_n = B \cap V_n \cap T_t^{-1}(V'_n),$$

so that $\cup_n B_n = B$ has full μ_0-measure.

If $x \in B_n$, it satisfies $x \in V_n$ and $T_t(x) \in V'_n$. From Inequality (1.5.3) above, we obtain

$$\begin{aligned}\psi(T_t(x)) - c_{t-s,L}(T_s(x), T_t(x)) &\leq \varphi_s^n(T_s(x)) \leq \varphi_s(T_s(x)) \\ &\leq \psi_s(T_s(x)) \leq \psi_s^n(T_s(x)) \leq \varphi(x) + c_{s,L}(x, T_s(x))\end{aligned}$$

Since $B_n \subset B$, for $x \in B_n$, we have $\psi(T_t(x)) - \varphi(x) = c_{t,L}(x, T_t(x))$. Combining this with Equality (1.5.2), we conclude that the two extreme terms in the inequality above are equal. Hence, for every $x \in B_n$, we have

$$\begin{aligned}\psi(T_t(x)) - c_{t-s,L}(T_s(x), T_t(x)) &= \varphi_s^n(T_s(x)) = \varphi_s(T_s(x)) \\ &= \psi_s(T_s(x)) = \psi_s^n(T_s(x)) = \varphi(x) + c_{s,L}(x, T_s(x)). \quad (1.5.5)\end{aligned}$$

In particular, we get

$$\forall x \in B, \quad \psi_s(T_s(x)) = \varphi(x) + c_{s,L}(x, T_s(x)),$$

or equivalently

$$\psi_s(y) - \varphi(x) = c_{s,L}(x, y) \quad \text{for } \gamma_s\text{-a.e. } (x, y).$$

Since (φ, ψ_s) is a (Borel) $c_{s,L}$-subsolution, it follows that the pair (φ, ψ_s) is $(c_{s,L}, \gamma_s)$-calibrated. Therefore, by Theorem 1.2.3 we get that $\gamma_s = (\mathrm{Id}_M \tilde{\times} T_s)_\# \mu_0$ is an optimal plan for $(c_{s,L}, \mu_0, \mu_s)$. Moreover, since $c_{s,L}$ is γ_s-integrable, the infimum in the Kantorovitch problem (1.2.1) in Theorem 1.3.2 with cost $c_{s,L}$ is finite, and therefore there exists a unique optimal transport plan. This proves (i).

Note for further reference that a similar argument, using the equality

$$\forall x \in B, \quad \psi(T_t(x)) = \varphi_s(T_s(x)) + c_{t-s,L}(T_s(x), T_t(x)),$$

which follows from Equation (1.5.5) above, shows that the measure $\hat{\gamma}_s = (T_s \tilde{\times} T_t)_\# \mu_0$ is an optimal plan for the cost $c_{t-s,L}$ and the pair of measures (μ_s, μ_t).

We now want to prove (ii). Since B is the increasing union of $B_n = B \cap V_n \cap T_t^{-1}(V_n')$, it suffices to prove that T_s is injective on B_n and that the restriction $T^{-1}|_{T(B_n)}$ is locally Lipschitz on $T_s(B_n)$.

Since $B_n \subset V_n$, by Inequality (1.5.3) above we have

$$\forall x \in B_n, \forall y \in M, \quad \varphi_s^n(y) \le \psi_s^n(y) \le \varphi(x) + c_{s,L}(x, y). \quad (1.5.6)$$

Moreover, by Equality (1.5.5) above

$$\forall x \in B_n, \quad \varphi_s^n(T_s(x)) = \psi_s^n(T_s(x)) = \varphi(x) + c_{s,L}(x, T_s(x)). \quad (1.5.7)$$

In particular, we have $\varphi_s^n \le \psi_s^n$ everywhere with equality at every point of $T_s(B_n)$. As we have said above, both functions ψ_s^n and $-\varphi_s^n$ are locally semi-concave with a linear modulus. It follows, from Theorem 6.1.19, that both derivatives $d_z\varphi_s^n, d_z\psi_s^n$ exist and are equal for $z \in T_s(B_n)$. Moreover, the map $z \mapsto d_z\varphi_s^n = d_z\psi_s^n$ is locally Lipschitz on $T_s(B_n)$. Note that we also get from (1.5.6) and (1.5.7) above that for a fixed $x \in B_n$, we have $\varphi_s^n \le \varphi(x) + c_{s,L}(x, \cdot)$ everywhere with equality at $T_s(x)$. Since φ_n is semi-convex, using that $c_{s,L}(x,\cdot)$ is semi-concave, again by Theorem 6.1.19, we obtain that the partial derivative $\dfrac{\partial c_{s,L}}{\partial y}(x, T_s(x))$ of $c_{s,L}$ with respect to the second variable exists and is equal to $d_{T_s(x)}\varphi_s^n = d_{T_s(x)}\psi_s^n$. Since $\gamma_x : [0,t] \to M$ is an L-minimizer with $\gamma_x(0) = x$ and $\gamma_x(s) = T_s(x)$, it follows from Corollary 6.2.20 that

$$d_{T_s(x)}\psi_s^n = \frac{\partial c_{s,L}}{\partial y}(x, T_s(x)) = \frac{\partial L}{\partial v}(\gamma_x(s), \dot\gamma_x(s)).$$

Since γ_x is an L-minimizer, its speed curve is an orbit of the Euler-Lagrange flow, and therefore

$$(T_s(x), d_{T_s(x)}\psi_s^n) = \mathscr{L}((\gamma_x(s), \dot\gamma_x(s)) = \mathscr{L}\phi_s^L(\gamma_x(0), \dot\gamma_x(0)),$$

and

$$x = \pi\phi_{-s}^L\mathscr{L}^{-1}(T_s(x), d_{T_s(x)}\psi_s^n).$$

It follows that T_s is injective on B_n with inverse given by the map $\theta_n : T_s(B_n) \to B_n$ defined, for $z \in T_s(B_n)$, by

$$\theta_n(z) = \pi\phi_{-s}^L\mathscr{L}^{-1}(z, d_z\psi_s^n).$$

Note that the map θ_n is locally Lipschitz on $T_s(B_n)$, since this is the case for $z \mapsto d_z\psi_s^n$, and both maps $\phi_{-s}^L, \mathscr{L}^{-1}$ are C^1, since L is a Tonelli Lagrangian. An analogous argument proves the countably Lipschitz regularity of $\hat T_s = T_t T_s^{-1}$ in part (iv). Finally the optimality of T_s^{-1} simply follows from

$$\begin{aligned}\min_{\gamma\in\Pi(\mu_s,\mu_0)}\left\{\int_{M\times M}\bar c_{s,L}(x,y)\,d\gamma(x,y)\right\} &= \min_{\gamma\in\Pi(\mu_0,\mu_s)}\left\{\int_{M\times M}c_{s,L}(x,y)\,d\gamma(x,y)\right\}\\ &= \int_M c_{s,L}(x, T_s(x))\,d\mu_0(x)\\ &= \int_M \bar c_{s,L}(y, T_s^{-1}(y))\,d\mu_s(y).\end{aligned}$$

Part (iii) of the Theorem follows from part (ii). In fact, if $A \subset M$ is Lebesgue negligible, the image $T_s^{-1}(T_s(B)\cap A)$ is also Lebesgue negligible, since T_s^{-1} is countably Lipschitz on $T_s(B)$, and therefore sends Lebesgue negligible subsets to Lebesgue negligible subsets. It remains to note, using that B is of full μ_0-measure, that $\mu_s(A) = T_{s\#}\mu_0(A) = \mu_0(T_s^{-1}(T_s(B)\cap A)) = 0$.

To prove part (iv), we already know that $\hat\gamma_s = (T_s \tilde\times T_t)_\#\mu_0$ is an optimal plan for the cost $c_{t-s,L}$ and the measures (μ_s, μ_t). Since the Borel set B is of full μ_0-measure, and $T_s : B \to T_s(B)$ is a bijective Borel measurable map, we obtain that T_s^{-1} is a Borel map, and $\mu_0 = T_{s\#}^{-1}\mu_s$. It follows that

$$\hat\gamma_s = (\mathrm{Id}_M \tilde\times T_t T_s^{-1})_\#\mu_s.$$

Therefore the composition $T_t T_s^{-1}$ is an optimal transport map for the cost $c_{t-s,L}$ and the pair of measures (μ_s, μ_t), and it is the unique one since $c_{t-s,L}$ is $\hat{\gamma}_s$-integrable and μ_s is absolutely continuous with respect to the Lebesgue measure. $\square$

Remark 1.5.3. We observe that, in proving the uniqueness statement in parts (i) and (iv) of the above theorem, we needed the full generality of Theorem 1.4.2, in which we only assume that the infimum in the Kantorovitch problem is finite. Indeed, assuming

$$\int_{M\times M} c_{t,L}(x,y)\, d\mu_0(x) d\mu_t(y) < +\infty,$$

there is a priori no reason for which the two integrals

$$\int_{M\times M} c_{s,L}(x,y)\, d\mu_0(x) d\mu_s(y), \qquad \int_{M\times M} c_{t-s,L}(x,y)\, d\mu_s(x) d\mu_t(y)$$

would have to be finite. So the existence and uniqueness of a transport map in Theorem 1.3.2 under the integrability assumption on c with respect to $\mu \otimes \nu$ instead of assumption (iv) would not have been enough to obtain Theorem 1.5.2.

Remark 1.5.4. We remark that, if both μ_0 and μ_t are not assumed to be absolutely continuous, and therefore no optimal transport map necessarily exists, one can still define an "optimal" interpolation $(\mu_s)_{0\leq s\leq t}$ between μ_0 and μ_t using some measurable selection theorem (see [133, Chapter 7]). Then, adapting our proof, one still obtains that, for any $s \in (0,t)$, there exists a unique optimal transport map S_s for $(\bar{c}_{s,L}, \mu_s, \mu_0)$ (resp. a unique optimal transport map $\hat{S}_s$ for $(c_{t-s,L}, \mu_s, \mu_t)$), and this map is countably Lipschitz.

We also observe that, if the manifold is compact, our proof shows that the above maps are globally Lipschitz (see [22]).

1.6 The Wasserstein space W_2

Let (M, g) be a smooth complete Riemannian manifold, equipped with its geodesic distance d and its volume measure vol. We denote with $P_2(M)$ the set of probability measures on M with finite 2-order moment, that is

$$\int_M d^2(x, x_0)\, d\mu(x) < +\infty \quad \text{for a certain } x_0 \in M.$$

We remark that, by the triangle inequality for d, the definition does not depends on the point x_0. The space $P_2(M)$ can be endowed of the so called Wasserstein distance W_2:

$$W_2^2(\mu_0, \mu_1) := \min_{\gamma \in \Pi(\mu_0,\mu_1)} \left\{ \int_{M\times M} d^2(x,y)\, d\gamma(x,y) \right\}.$$

The quantity W_2 will be called the *Wasserstein distance of order 2* between μ_0 and μ_1. It is well-known that it defines a finite metric on $P_2(M)$, and so one can speak about geodesic in the metric space (P_2, W_2). This space turns out, indeed, to be a length space (see for example [132], [133]). We denote with $P_2^{ac}(M)$ the subset of $P_2(M)$ that consists of the Borel probability measures on M that are absolutely continuous with respect to vol.

By all the result proved before, it is simple to prove the following:

Proposition 1.6.1. *$P_2^{ac}(M)$ is a geodesically convex subset of $P_2(M)$. Moreover, if $\mu_0, \mu_1 \in P_2^{ac}(M)$, then there is a unique Wasserstein geodesic $\{\mu_t\}_{t\in[0,1]}$ joining μ_0 to μ_1, which is given by*

$$\mu_t = (T_t)_\sharp \mu_0 := (\exp[t\tilde{\nabla}\varphi])_\sharp \mu_0,$$

where $T(x) = \exp_x[\tilde{\nabla}_x\varphi]$ is the unique transport map from μ_0 to μ_1 which is optimal for the cost $\frac{1}{2}d^2(x,y)$ (and so also optimal for the cost $d^2(x,y)$). Moreover:

1. *T_t is the unique optimal transport map from μ_0 to μ_t for all $t \in [0,1]$;*
2. *T_t^{-1} the unique optimal transport map from μ_t to μ_0 for all $t \in [0,1]$ (and, if $t \in [0,1)$, it is locally Lipschitz);*
3. *$T \circ T_t^{-1}$ the unique optimal transport map from μ_t to μ_1 for all $t \in [0,1]$ (and, if $t \in (0,1]$, it is locally Lipschitz).*

Since we know that the transport is unique, the proof is quite standard. However, for completeness, we give all the details.

Proof. Let $\{\mu_t\}_{t\in[0,1]}$ be a Wasserstein geodesic joining μ_0 to μ_1. Fix $t \in (0,1)$, and let γ_t (resp. $\hat{\gamma}_t$) be an optimal transport plan between μ_0 and μ_t (resp. μ_t and μ_1) (in effect, we know that γ_t is a graph and it is unique, but we will not use this fact). We now define the probability measure on $M \times M \times M$

$$\lambda_t(dx, dy, dz) := \int_M \gamma_t(dx|y) \times \hat{\gamma}_t(dz|y)\, d\mu_t(y),$$

where $\gamma_t(dx,dy) = \int_M \gamma_t(dx|y)\, d\mu_t(y)$ and $\hat{\gamma}_t(dy,dz) = \int_M \hat{\gamma}_t(dz|y)\, d\mu_t(y)$ are the disintegrations of γ_t and $\hat{\gamma}_t$ with respect to μ_t. Then, if we define

$$\tilde{\gamma}_t := \pi_\sharp^{1,3}\lambda_t,$$

it is simple to check that $\tilde{\gamma}_t$ is a transport plan from μ_0 to μ_1. Now, since $\{\mu_t\}_{t\in[0,1]}$ is a geodesic, we have that

$$\begin{aligned} W_2(\mu_0,\mu_1) &\le \|d(x,z)\|_{L^2(\tilde{\gamma}_t, M\times M)} = \|d(x,z)\|_{L^2(\lambda_t, M\times M\times M)} \\ &\le \|d(x,y)\|_{L^2(\lambda_t, M\times M\times M)} + \|d(y,z)\|_{L^2(\lambda_t, M\times M\times M)} \\ &= \|d(x,y)\|_{L^2(\gamma_t, M\times M)} + \|d(y,z)\|_{L^2(\hat{\gamma}_t, M\times M)} \\ &= W_2(\mu_0,\mu_t) + W_2(\mu_t,\mu_1) = W_2(\mu_0,\mu_1). \end{aligned} \tag{1.6.1}$$

This proves that $\tilde{\gamma}_t$ is an optimal transport plan between μ_0 and μ_1, which implies that $\tilde{\gamma}_t$ is supported on the graph of T. Moreover, since in (1.6.1) all the inequalities are indeed equalities, we get that

$$d(x,z) = d(x,y) + d(y,z) \quad \text{for } \lambda_t\text{-a.e. } (x,y,z) \in M \times M \times M$$

that is, y is on a geodesic from x to z. Moreover, since $W_2(\mu_0, \mu_t) = tW_2(\mu_0, \mu_1)$, we also have

$$d(x,y) = td(x,z), \quad d(y,z) = (1-t)d(x,z) \quad \text{for } \lambda_t\text{-a.e. } (x,y,z) \in M \times M \times M.$$

Since, by Remark 1.5.1, the geodesic from x to $T(x)$ is unique for μ_0-a.e. x, we conclude that λ is concentrated on the subset $\{(x, T_t(x), T(x))\}_{x\in\operatorname{supp}(\mu_0)}$, which implies that $\mu_t = (T_t)_\sharp\mu_0$. Moreover we see that $\mu_t := (T_t)_\sharp\mu_0 \in P_2^{ac}(M)$. In fact,

$$\begin{aligned}
\int_M d^2(x,x_0)\, d\mu_t(x) &= \int_M d^2(T_t(x), x_0)\, d\mu_0(x) \\
&\le 2\int_M \left[d^2(x,x_0) + d^2(x, T_t(x))\right]\, d\mu_0(x) \\
&\le 2\int_M \left[d^2(x,x_0) + d^2(x, T(x))\right]\, d\mu_0(x) \\
&\le 4\int_M \left[d^2(x,x_0) + d^2(x_0, T(x))\right]\, d\mu_0(x) \\
&= 4\int_M d^2(x,x_0)\, d\mu_0(x) + 4\int_M d^2(x_0,y)\, d\mu_1(y) < +\infty,
\end{aligned}$$

and the result in Section 1.5 tells us that μ_t is absolutely continuous. Using the notation of Section 1.4, we have

$$c_t(x,y) = \inf_{\gamma(0)=x,\ \gamma(t)=y} \int_0^t \frac{1}{2}\|\dot{\gamma}(s)\|^2_{\gamma(s)}\, ds = \frac{1}{2t}d^2(x,y).$$

Since T_t and T_t^{-1} are optimal for the cost function $\frac{1}{2t}d^2(x,y)$, and $T \circ T_t^{-1}$ is optimal for the cost function $\frac{1}{2(1-t)}d^2(x,y)$, we get that T_t, T_t^{-1} and $T \circ T_t^{-1}$ are optimal also for the cost $d^2(x,y)$. □

The above result tells us that also $(P_2^{ac}(M), W_2)$ is a length space.

1.6.1 Regularity, concavity estimate and a displacement convexity result

We now consider the cost function $c(x,y) = \frac{1}{2}d^2(x,y)$. Let $\mu, \nu \in P^{ac}(M)$ with $W_2(\mu,\nu) < +\infty$, and let us denote with f and g their respective densities with respect to vol. Let

$$T(x) = \exp_x[\tilde{\nabla}_x\varphi]$$

be the unique optimal transport map from μ to ν.

We recall that locally semi-convex (or semi-concave) functions with linear modulus admit vol-a.e. a second order Taylor expansion (see [16], [50]). Let us recall the definition of approximate hessian.

Definition 1.6.2 (approximate hessian). We say that $f : M \to \mathbb{R}^m$ has a *approximate hessian* at $x \in M$ if there exists a function $h : M \to \mathbb{R}$ such that the set $\{f = h\}$ has density 1 at x with respect to the Lebesgue measure and h admits a second order Taylor expansion at x, that is, there exists a self-adjoint operator $H : T_xM \to T_xM$ such that

$$h(\exp_x w) = h(x) + \langle \nabla_x h, w\rangle + \frac{1}{2}\langle Hw, w\rangle + o(\|w\|_x^2).$$

In this case the approximate hessian is defined as $\tilde{\nabla}_x^2 f := H$.

As in the case of the approximate differential, it is not difficult to show that this definition makes sense.

Observing that $d^2(x, y)$ is locally semi-concave with linear modulus (see [66, Appendix]), we get that φ_n is locally semi-convex with linear modulus for each n. Thus we can define μ-a.e. an approximate hessian for φ (see Definition 1.6.2):

$$\tilde{\nabla}_x^2 \varphi := \nabla_x^2 \varphi_n \quad \text{for } x \in A_n \cap E_n,$$

where A_n was defined in the proof of Complement 1.3.4, E_n denotes the full μ-measure set of points where φ_n admits a second order Taylor expansion, and $\nabla_x^2\varphi_n$ denotes the self-adjoint operator on T_xM that appears in the Taylor expansion on φ_n at x. Let us now consider, for each set $F_n := A_n \cap E_n$, an increasing sequence of compact sets $K_m^n \subset F_n$ such that $\mu(F_n \setminus \cup_m K_m^n) = 0$. We now define the measures $\mu_m^n := \mu \llcorner K_m^n$ and $\nu_m^n := T_\sharp \mu_m^n = (\exp[\nabla\varphi_n])_\sharp \mu_m^n$, and we renormalize them in order to obtain two probability measures:

$$\hat{\mu}_m^n := \frac{\mu_m^n}{\mu_m^n(M)} \in P_2^{ac}(M), \quad \hat{\nu}_m^n := \frac{\nu_m^n}{\nu_m^n(M)} = \frac{\nu_m^n}{\mu_m^n(M)} \in P_2^{ac}(M).$$

We now observe that T is still optimal. In fact, if this were not the case, we would have

$$\int_{M\times M} c(x, S(x))\, d\hat{\mu}_m^n(x) < \int_{M\times M} c(x, T(x))\, d\hat{\mu}_m^n(x)$$

for a certain S transport map from $\hat{\mu}_m^n$ to $\hat{\nu}_m^n$. This would imply that

$$\int_{M\times M} c(x, S(x))\, d\mu_m^n(x) < \int_{M\times M} c(x, T(x))\, d\mu_m^n(x),$$

and so the transport map

$$\tilde{S}(x) := \begin{cases} S(x) & \text{if } x \in K_m^n, \\ T(x) & \text{if } x \in M \setminus K_m^n \end{cases}$$

would have a cost strictly less than the cost of T, which would contradict the optimality of T.

We will now apply the results of [50] to the compactly supported measures $\hat{\mu}_m^n$ and $\hat{\nu}_m^n$ in order to get information on the transport problem from μ to ν. In what follows we will denote by $\nabla_x d_y^2$ and by $\nabla_x^2 d_y^2$, respectively, the gradient and the hessian with respect to x of $d^2(x,y)$, and by $d_x \exp$ and $d(\exp_x)_v$ the two components of the differential of the map $TM \ni (x,v) \mapsto \exp_x[v] \in M$ (whenever they exist). By [50, Theorem 4.2], we get the following.

Theorem 1.6.3 (Jacobian identity a.e.). *There exists a subset $E \subset M$ such that $\mu(E) = 1$ and, for each $x \in E$, $Y(x) := d(\exp_x)_{\tilde{\nabla}_x \varphi}$ and $H(x) := \frac{1}{2}\nabla_x^2 d_{T(x)}^2$ both exist and we have*

$$f(x) = g(T(x)) \det[Y(x)(H(x) + \tilde{\nabla}_x^2 \varphi)] \neq 0.$$

Proof. It suffices to observe that [50, Theorem 4.2] applied to $\hat{\mu}_m^n$ and $\hat{\nu}_m^n$ gives that, for μ-a.e. $x \in K_m^n$,

$$\frac{f(x)}{\mu_m^n(M)} = \frac{g(T(x))}{\mu_m^n(M)} \det[Y(x)(H(x) + \nabla_x^2 \varphi_n)] \neq 0,$$

which implies

$$f(x) = g(T(x)) \det[Y(x)(H(x) + \tilde{\nabla}_x^2 \varphi)] \neq 0 \quad \text{for } \mu\text{-a.e. } x \in K_m^n.$$

Passing to the limit as $m, n \to +\infty$ we get the result. □

We can thus define μ-a.e. the (weak) differential of the transport map at x as

$$d_x T := Y(x)\big(H(x) + \tilde{\nabla}_x^2 \varphi\big).$$

Let us prove now that, indeed, $T(x)$ is approximately differentiable μ-a.e., and that the above differential coincides with the approximate differential of T. In order to prove this fact, let us first make a formal computation. Observe that since the map $x \mapsto \exp_x[-\frac{1}{2}\nabla_x d_y^2] = y$ is constant, we have

$$0 = d_x(\exp_x[-\tfrac{1}{2}\nabla_x d_y^2]) = d_x \exp[-\tfrac{1}{2}\nabla_x d_y^2] - d(\exp_x)_{-\frac{1}{2}\nabla_x d_y^2}\big(\tfrac{1}{2}\nabla_x^2 d_y^2\big) \quad \forall y \in M,$$

By differentiating (in the approximate sense) the equality $T(x) = \exp[\tilde{\nabla}_x \varphi]$ and recalling the equality $\tilde{\nabla}_x \varphi = -\frac{1}{2}\nabla_x d^2_{T(x)}$, we obtain

$$\begin{aligned}\tilde{d}_x T &= d(\exp_x)_{\tilde{\nabla}_x \varphi}\big(\tilde{\nabla}^2_x \varphi\big) + d_x \exp[\tilde{\nabla}_x \varphi] \\ &= d(\exp_x)_{\tilde{\nabla}_x \varphi}\big(\tilde{\nabla}^2_x \varphi\big) + d(\exp_x)_{-\frac{1}{2}\nabla_x d^2_{T(x)}}\big(\tfrac{1}{2}\nabla^2_x d^2_{T(x)}\big) \\ &= d(\exp_x)_{\tilde{\nabla}_x \varphi}\big(H(x) + \tilde{\nabla}^2_x \varphi\big),\end{aligned}$$

as wanted. In order to make the above proof rigorous, it suffices to observe that for μ-a.e. x, $T(x) \notin cut(x)$, where $cut(x)$ is defined as the set of points $z \in M$ which cannot be linked to x by an extendable minimizing geodesic. Indeed we recall that the square of the distance fails to be semi-convex at the cut locus, that is, if $x \in cut(y)$, then

$$\inf_{0<\|v\|_x<1} \frac{d^2_y(\exp_x[v]) - 2d^2_y(x) + d^2_y(\exp_x[-v])}{|v|^2} = -\infty$$

(see [50, Proposition 2.5]). Now fix $x \in F_n$. Since we know that $\frac{1}{2}d^2(z, T(x)) \geq \psi(T(x)) - \varphi_n(z)$ with equality for $z = x$, we obtain a bound from below of the hessian of $d^2_{T(x)}$ at x in terms of the hessian of φ_n at x (see the proof of [50, Proposition 4.1(a)]). Thus, since each φ_n admits vol-a.e. a second order Taylor expansion, we obtain that, for μ-a.e. x,

$$x \notin cut(T(x)), \quad \text{or equivalently} \quad T(x) \notin cut(x).$$

This implies that all the computations we made above in order to prove the formula for $\tilde{d}_x T$ are correct. Indeed the exponential map $(x, v) \mapsto \exp_x[v]$ is smooth if $\exp_x[v] \notin cut(x)$, the function d^2_y is smooth around any $x \notin cut(y)$ (see [50, Paragraph 2]), and $\tilde{\nabla}_x \varphi$ is approximatively differentiable μ-a.e. Thus, recalling that, once we consider the right composition of an approximatively differentiable map with a smooth map, the standard chain rule holds (see the remarks after Definition 1.3.3), we have proved the following regularity result for the transport map.

Proposition 1.6.4 (approximate differentiability of the transport map). *The transport map is approximatively differentiable for μ-a.e. x, and its approximate differential is given by the formula*

$$\tilde{d}_x T = Y(x)\big(H(x) + \tilde{\nabla}^2_x \varphi\big),$$

where Y and H are defined in Theorem 1.6.3.

To prove our displacement convexity result, the following change of variables formula will be useful.

Proposition 1.6.5 (change of variables for optimal maps). *If $A : [0 + \infty) \to \mathbb{R}$ is a Borel function such that $A(0) = 0$, then*

$$\int_M A(g(y))\, d\,\mathrm{vol}(y) = \int_E A\left(\frac{f(x)}{J(x)}\right) J(x)\, d\,\mathrm{vol}(x),$$

where $J(x) := \det[Y(x)(H(x) + \tilde{\nabla}_x^2 \varphi)] = \det[\tilde{d}_x T]$ (either both integrals are undefined or both take the same value in $\overline{\mathbb{R}}$).

The proof follows by the Jacobian identity proved in Theorem 1.6.3, exactly as in [50, Corollary 4.7].

Let us now define for $t \in [0,1]$ the measure $\mu_t := (T_t)_\sharp \mu$, where

$$T_t(x) = \exp_x[t\tilde{\nabla}_x \varphi].$$

By the results in [66] and Proposition 1.6.1, we know that T_t coincides with the unique optimal map pushing μ forward to μ_t, and that μ_t is absolutely continuous with respect to vol for any $t \in [0,1]$.

Given $x, y \in M$, following [50], we define for $t \in [0,1]$

$$Z_t(x,y) := \{z \in M \mid d(x,z) = td(x,y) \text{ and } d(z,y) = (1-t)d(x,y)\}.$$

If N is now a subset of M, we set

$$Z_t(x,N) := \cup_{y \in N} Z_t(x,y).$$

Letting $B_r(y) \subset M$ denote the open ball of radius $r > 0$ centered at $y \in M$, for $t \in (0,1]$ we define

$$v_t(x,y) := \lim_{r \to 0} \frac{\mathrm{vol}(Z_t(x, B_r(y)))}{\mathrm{vol}(B_{tr}(y))} > 0$$

(the above limit always exists, though it will be infinite when x and y are conjugate points; see [50]). Arguing as in the proof of Theorem 1.6.3, by [50, Lemma 6.1] we get the following.

Theorem 1.6.6 (Jacobian inequality). *Let E be the set of full μ-measure given by Theorem 1.6.3. Then for each $x \in E$, $Y_t(x) := d(\exp_x)_{t\tilde{\nabla}_x \varphi}$ and $H_t(x) := \frac{1}{2}\nabla_x^2 d^2_{T_t(x)}$ both exist for all $t \in [0,1]$ and the Jacobian determinant*

$$J_t(x) := \det[Y_t(x)(H_t(x) + t\tilde{\nabla}_x^2 \varphi)] \tag{1.6.2}$$

satisfies

$$J_t^{\frac{1}{n}}(x) \geq (1-t)\,[v_{1-t}(T(x),x)]^{\frac{1}{n}} + t\,[v_t(x,T(x))]^{\frac{1}{n}}\, J_1^{\frac{1}{n}}(x).$$

We now consider as source measure $\mu_0 = \rho_0\, d\,\mathrm{vol}(x) \in P^{ac}(M)$ and as target measure $\mu_1 = \rho_1\, d\,\mathrm{vol}(x) \in P^{ac}(M)$, and we assume as before that $W_2(\mu_0,\mu_1) < +\infty$. By Proposition 1.6.1 we have

$$\mu_t = (T_t)_\sharp[\rho_0\, d\,\mathrm{vol}] = \rho_t\, d\,\mathrm{vol} \in P_2^{ac}(M)$$

for a certain $\rho_t \in L^1(M, d\,\mathrm{vol})$.

We now want to consider the behavior of the functional

$$U(\rho) := \int_M A(\rho(x))\, d\,\mathrm{vol}(x)$$

along the path $t \mapsto \rho_t$. In Euclidean spaces, this path is called *displacement interpolation* and the functional U is said to be *displacement convex* if

$$[0,1] \ni t \mapsto U(\rho_t) \quad \text{is convex for every } \rho_0, \rho_1.$$

A sufficient condition for the displacement convexity of U in $\mathbb{R}^n$ is that $A : [0,+\infty) \to \mathbb{R}\cup\{+\infty\}$ satisfy

$$(0,+\infty) \in s \mapsto s^n A(s^{-n}) \text{ is convex and nonincreasing, with } A(0) = 0 \tag{1.6.3}$$

(see [106], [108]). Typical examples include the entropy $A(\rho) = \rho\log\rho$ and the L^q-norm $A(\rho) = \frac{1}{q-1}\rho^q$ for $q \geq \frac{n-1}{n}$.

By all the results collected above, arguing as in the proof of [50, Theorem 6.2], we can prove that the displacement convexity of U is still true on Ricci nonnegative manifolds under the assumption (1.6.3).

Theorem 1.6.7 (displacement convexity on Ricci nonnegative manifolds). *If* $\mathrm{Ric} \geq 0$ *and A satisfies* (1.6.3)*, then U is displacement convex.*

Proof. As we remarked above, T_t is the optimal transport map from μ_0 to μ_t. So, by Theorem 1.6.3 and Proposition 1.6.5, we get

$$U(\rho_t) = \int_M A(\rho_t(x))\, d\,\mathrm{vol}(x) = \int_{E_t} A\left(\frac{\rho_0(x)}{\left(J_t^{\frac{1}{n}}(x)\right)^n}\right)\left(J_t^{\frac{1}{n}}(x)\right)^n d\,\mathrm{vol}(x), \tag{1.6.4}$$

where E_t is the set of full μ_0-measure given by Theorem 1.6.3 and $J_t(x) \neq 0$ is defined in (1.6.2). Since $\mathrm{Ric} \geq 0$, we know that $v_t(x,y) \geq 1$ for every $x, y \in M$ (see [50, Corollary 2.2]). Thus, for fixed $x \in E_1$, Theorem 1.6.6 yields the concavity of the map

$$[0,1] \ni t \mapsto J_t^{\frac{1}{n}}(x).$$

Composing this function with the convex nonincreasing function $s \mapsto s^n A(s^{-n})$ we get the convexity of the integrand in (1.6.4). The only problem we run into in trying to conclude the displacement convexity of U is that the domain of integration appears to depend on t. But, since by Theorem 1.6.3 E_t is a set of full μ_0-measure for any $t \in [0,1]$, we obtain that, for fixed $t, t', s \in [0,1]$,

$$U(\rho_{(1-s)t+st'}) \leq (1-s)U(\rho_t) + sU(\rho_{t'}),$$

simply by computing each of the three integrals above on the full measure set $E_t \cap E_{t'} \cap E_{(1-s)t+st'}$. □

1.7 Displacement convexity on Riemannian manifolds

For the past few years, there has been ongoing research to study the links between Riemannian geometry and optimal transport of measures [132, 133]. In particular, it was recently found that lower bounds on the Ricci curvature tensor can be recast in terms of convexity properties of certain nonlinear functionals defined on spaces of probability measures [50, 97, 116, 126, 127, 128]. In this section we solve a natural problem in this field by establishing the equivalence of several such formulations.

Before explaining our results in more detail, let us give some notation and background. Let (M, g) be a smooth complete connected n-dimensional Riemannian manifold, equipped with its geodesic distance d and its volume measure vol. Let $P(M)$ be the set of probability measures on M. For any real number $p \geq 1$, we denote by $P_p(M)$ the set of probability measures μ such that

$$\int_M d^p(x, x_0)\, d\mu(x) < \infty \qquad \text{for some } x_0 \in M.$$

The set $P_2(M)$ is equipped with the Wasserstein distance of order 2, denoted by W_2: This is the square root of the optimal transport cost functional, when the cost function $c(x,y)$ coincides with the squared distance $d^2(x,y)$; see for instance [133, Definition 6.1]. Then $P_2(M)$ is a metric space, and even a length space; that is, any two probability measures in $P_2(M)$ are joined by at least one geodesic curve $(\mu_t)_{0 \leq t \leq 1}$. (Here and in the sequel, by convention geodesics are supposed to be globally minimizing and to have constant speed.)

A basic representation theorem (see [97, Proposition 2.10] or [133, Corollary 7.22]) states that any Wasserstein geodesic curve necessarily takes the form $\mu_t = (e_t)_* \Pi$, where Π is a probability measure on the set Γ of minimizing geodesics $[0,1] \to M$, the symbol

$*$ stands for push-forward, and $e_t : \Gamma \to M$ is the evaluation at time t: $e_t(\gamma) := \gamma(t)$. So the optimal transport problem between two probability measures μ_0 and μ_1 produces three related objects:

- an optimal coupling π of μ_0 and μ_1, which is a probability measure on $M \times M$ whose marginals are μ_0 and μ_1, achieving the lowest possible cost for the transport between these measures;

- a path $(\mu_t)_{0 \le t \le 1}$ in the space of probability measures;

- a probability measure Π on the space of geodesics, such that $(e_t)_*\Pi = \mu_t$ and $(e_0, e_1)_*\Pi = \pi$. Such a Π is called a dynamical optimal transference plan [133, Definition 7.20].

The core of the studies in [50, 97, 116, 126, 127, 128] lies in the analysis of the convexity properties of certain nonlinear functionals along geodesics in $P_2(M)$, defined below:

Definition 1.7.1 (Nonlinear functionals of probability measures). Let ν be a reference measure on M, absolutely continuous with respect to the volume measure. Let $U : \mathbb{R}_+ \to \mathbb{R}$ be a continuous convex function with $U(0) = 0$; let $U'(\infty)$ be the limit of $U(r)/r$ as $r \to \infty$. Let μ be a probability measure on M and let $\mu = \rho\nu + \mu_s$ be its Lebesgue decomposition with respect to ν.

(i) If $U(\rho)$ is bounded below by a ν-integrable function, then the quantity $U_\nu(\mu)$ is defined by the formula

$$U_\nu(\mu) = \int_M U(\rho(x))\,\nu(dx) + U'(\infty)\,\mu_s[M].$$

(ii) If π is a probability measure on $M \times M$, admitting μ as first marginal, β is a positive function on $M \times M$, and $\beta\, U(\rho/\beta)$ is bounded below (as a function of x, y) by a ν-integrable function of x, then the quantity $U^\beta_{\pi,\nu}(\mu)$ is defined by the formula

$$U^\beta_{\pi,\nu}(\mu) = \int_{M \times M} U\left(\frac{\rho(x)}{\beta(x,y)}\right)\,\beta(x,y)\pi(dy|x)\,\nu(dx) + U'(\infty)\,\mu_s[M],$$

where $\pi(dy|x)$ is the disintegration of $\pi(dx\,dy)$ with respect to the x variable.

Remark 1.7.2. Sufficient conditions for U_ν and $U^\beta_{\pi,\nu}$ to be well-defined are discussed in [133, Theorems 17.8 and 17.28, Application 17.29] and will not be addressed here.

Remark 1.7.3. If $U'(\infty) = \infty$, then finiteness of $U_\nu(\mu)$ implies that μ is absolutely continuous with respect to ν. This is not true if $U'(\infty) < \infty$.

The various notions of convexity that are considered in [97, 126, 127, 128] belong to the following ones:

Definition 1.7.4 (Convexity properties). (i) Let U and ν be as in Definition 1.7.1, and let $\lambda \in \mathbb{R}$. We say that the functional U_ν is *λ-displacement convex* if for all Wasserstein geodesics $(\mu_t)_{0\le t\le 1}$ whose image lies in the domain of U_ν,

$$U_\nu(\mu_t) \le (1-t)\, U_\nu(\mu_0) + t\, U_\nu(\mu_1) - \frac{1}{2}\,\lambda\, t(1-t) W_2^2(\mu_0,\mu_1), \qquad \forall t \in [0,1]. \tag{1.7.1}$$

We say that the functional U_ν is *displacement convex with distortion β* if for all Wasserstein geodesics $(\mu_t)_{0\le t\le 1}$ whose image lies in the domain of U_ν, if $\pi(dx\,dy)$ stands for the associated optimal coupling between μ_0 and μ_1, and $\check{\pi}$ is obtained from π by exchanging the two variables x and y, then

$$U_\nu(\mu_t) \le (1-t)\, U^\beta_{\pi,\nu}(\mu_0) + t\, U^\beta_{\check{\pi},\nu}(\mu_1), \qquad \forall t \in [0,1]. \tag{1.7.2}$$

(ii) We say that U_ν is *weakly λ-displacement convex* (resp. *weakly displacement convex with distortion β*) if for all μ_0, μ_1 in the domain of U_ν, there is some Wasserstein geodesic from μ_0 to μ_1 along which (1.7.1) (resp. (1.7.2)) is satisfied.

(iii) We say that U_ν is *weakly λ-a.c.c.s. displacement convex* (resp. *weakly a.c.c.s. displacement convex with distortion β*) if condition (1.7.1) (resp. (1.7.2)) is satisfied along some Wasserstein geodesic when we further assume that μ_0, μ_1 are absolutely continuous and compactly supported.

Remark 1.7.5. The Wasserstein geodesic in (ii) and (iii) above is implicitly assumed to have its image entirely contained in the domain of the functional U_ν.

Remark 1.7.6. If U_ν is a λ-displacement convex functional, then the function $t \mapsto U_\nu(\mu_t)$ is λ-convex on $[0,1]$, i.e. for all $0 \le s \le s' \le 1$ and $t \in [0,1]$,

$$U_\nu(\mu_{(1-t)s+ts'}) \le (1-t)U_\nu(\mu_s) + tU_\nu(\mu_{s'}) - \frac{1}{2}\lambda t(1-t)(s'-s)^2 W_2(\mu_0,\mu_1)^2. \tag{1.7.3}$$

This is not *a priori* the case if we only assume that U_ν is weakly λ-displacement convex.

In short, *weakly* means that we require a condition to hold only for *some* geodesic between two measures, as opposed to *all* geodesics, and *a.c.c.s.* means that we only require the condition to hold when the two measures are absolutely continuous and compactly supported.

There are obvious implications (with or without distorsion)

$$\begin{gathered}\lambda\text{-displacement convex}\\ \Downarrow\\ \text{weakly } \lambda\text{-displacement convex}\\ \Downarrow\\ \text{weakly } \lambda\text{-a.c.c.s. displacement convex.}\end{gathered}$$

Although the natural convexity condition is arguably the one appearing in (i), that is, holding true along all Wasserstein geodesics, this condition is quite more delicate to study than the weaker conditions appearing in (ii) and (iii), in particular for stability issues: See [97, 126, 127]. In the same references the equivalence between (ii) and (iii) was established, at least for compact spaces [97, Proposition 3.21]. But the implication (ii) $\Rightarrow$ (i) remained open (and was listed as an open problem in a preliminary version of [133]). Here we shall fill this gap (at least for the functionals defined above), thus answering a natural question about the notion of displacement convexity. Here is our main result:

Theorem 1.7.7. *Let U, ν and β be as in Definition 1.7.1. Assume that U is Lipschitz. For each $a > 0$, define $U_a(r) = U(ar)/a$. Then*

(i) If $(U_a)_\nu$ is weakly λ-a.c.c.s. displacement convex for any $a \in (0,1]$, then U_ν is λ-displacement convex;

(ii) If $(U_a)_\nu$ is weakly a.c.c.s. displacement convex with distortion β for any $a \in (0,1]$, then U_ν is displacement convex with distortion β.

Among the consequences of Theorem 1.7.7 is the following corollary:

Corollary 1.7.8. *Let M be a smooth complete Riemannian manifold with nonnegative Ricci curvature and dimension n. Let $U(r) = -r^{1-1/n}$, and let ν be the volume measure on M. Then U_ν is displacement convex on $P_p(M)$, where $p = 2$ if $n \geq 3$, and p is any real number greater than 2 if $n = 2$.*

More generally, Theorem 1.7.7 makes it possible to drop the "weakly" in all displacement convexity characterizations of Ricci curvature bounds.

Before turning to the proof of Theorem 1.7.7, let us explain a bit more about the difficulties and the strategy of proof. Obviously, there are two problems to tackle: first, the possibility that μ_0 and/or μ_1 do not have compact support; and secondly, the possibility that μ_0 and/or μ_1 are singular with respect to the volume measure.

It was shown in [97, 126, 127] that inequalities such as (1.7.1) or (1.7.2) are *stable* under (weak) convergence. Then it is natural to approximate μ_0, μ_1 by compactly supported, absolutely continuous measures, and pass to the limit. This scheme of proof is enough to show the implication (iii) $\Rightarrow$ (ii) in Definition 1.7.4, but does not guarantee that we can attain *all* Wasserstein geodesics in this way — unless of course we know that there is a unique Wasserstein geodesic between μ_0 and μ_1.

To treat the difficulty arising from the possible non-compactness, we use the results of the previous sections, showing that the Wasserstein geodesic between any two absolutely continuous probability measures on a Riemannian manifold M is unique, even if they are not compactly supported.

The difficulty arising from the possible singularity of μ_0, μ_1 is less simple. If μ_0 and μ_1 are both singular, then there are in general several Wasserstein geodesics joining them. A most simple example is constructed by taking $\mu_0 = \delta_{x_0}$ and $\mu_1 = \delta_{x_1}$, where δ_x stands for the Dirac mass at x, and x_0, x_1 are joined by multiple geodesics. So it is part of the problem to regularize μ_0, μ_1 into absolutely continuous measures $\mu_{0,k}$, $\mu_{1,k}$ so that, as $k \to \infty$, the optimal transport between $\mu_{0,k}$ and $\mu_{1,k}$ converges to a *given* optimal transport between μ_0 and μ_1.

We handle this by a rather nonstandard regularization procedure, which roughly goes as follows. We start from a given dynamical optimal transference plan Π between μ_0 and μ_1, leave intact that part $\Pi^{(a)}$ of Π which corresponds to the absolutely continuous part of μ_0. Then we let displacement occur for a very short time at the level of that part $\Pi^{(s)}$ of Π corresponding to the singular part of μ_0. Next we regularize the resulting contribution of $\Pi^{(s)}$.

Let us illustrate this in the most basic case when $\mu_0 = \delta_{x_0}$ and $\mu_1 = \delta_{x_1}$. Let $\gamma = (\gamma_t)_{0\le t\le 1}$ be a given geodesic between x_0 and x_1; we wish to approximate the Wasserstein geodesic $(\delta_{\gamma_t})_{0\le t\le 1}$. Instead of directly regularizing μ_0 and μ_1, we shall first replace μ_0 by $\mu_\tau = \delta_{\gamma_\tau}$, where τ is positive but very small; and then regularize δ_{γ_τ} and δ_{x_1} into probability measures $\mu_{\tau,\varphi}$ and $\mu_{1,\varphi}$. What we have gained is that the geodesic joining γ_τ to $x_1 = \gamma_1$ is unique, so we may let $\tau \to 0$ and $\varphi \to 0$ in such a way that the Wasserstein geodesic joining $\mu_{\tau,\varphi}$ to $\mu_{1,\varphi}$ does converge to $(\delta_{\gamma_t})_{0\le t\le 1}$.

In a more general context, the procedure will be more tricky, and what will make it work is the following important property [133, Theorem 7.29]: *Geodesics in dynamical optimal transport plans do not cross at intermediate times.* In fact, if Π is a given dynamical optimal transport plan, then for each $t \in (0,1)$ one can define a measurable map $F_t : M \to \Gamma$ by the requirement that $F_t \circ e_t = \mathrm{Id}$, Π-almost surely. In understandable words, if γ is a geodesic along which there is optimal transport, then the position of γ at time t determines the whole geodesic γ. This property will ensure that $\Pi^{(a)}$ and $\Pi^{(s)}$ "do not overlap at intermediate times".

Finally, we note that the results in this section can be extended to more general situations outside the category of Riemannian manifolds: It is sufficient that the optimal transport between any two absolutely continuous probability measures be unique. In fact, there is a more general framework where these results still hold true, namely the case of *nonbranching* locally compact, complete length spaces. This extension is established, by a slightly different approach, in [133, Chapter 30].

1.7.1 Proofs

In the sequel, we shall use the notation $U_{a,\nu}$ for $(U_a)_\nu$. An important ingredient in the proof of Theorem 1.7.7 will be the following lemma, which has interest on its own (and will be used for different purposes in [133, Chapter 30]).

Lemma 1.7.9. *Let U be a Lipschitz convex function with $U(0) = 0$. Let μ_1, μ_2 be any two probability measures on M, and let Z_1, Z_2 be two positive numbers with $Z_1 + Z_2 = 1$. Then*

(i) $U_\nu(Z_1\mu_1 + Z_2\mu_2) \geq Z_1\, U_{Z_1,\nu}(\mu_1) + Z_2\, U_{Z_2,\nu}(\mu_2)$, with equality if μ_1 and μ_2 are singular to each other;

(ii) Let π_1, π_2 be two probability measures on $M \times M$, and let β be a positive measurable function on $M \times M$. Then

$$U^\beta_{Z_1\pi_1+Z_2\pi_2,\nu}(Z_1\mu_1 + Z_2\mu_2) \geq Z_1\, U^\beta_{Z_1,\pi_1,\nu}(\mu_1) + Z_2\, U^\beta_{Z_2,\pi_2,\nu}(\mu_2),$$

with equality if μ_1 and μ_2 are singular to each other.

Proof of Lemma 1.7.9. We start by the following remark: If x, y are nonnegative numbers, then

$$U(x + y) \geq U(x) + U(y). \tag{1.7.4}$$

Inequality (1.7.4) follows at once from the fact that $U(t)/t$ is a nondecreasing function of t, and thus

$$\frac{U(x)}{x} \leq \frac{U(x+y)}{x+y},\ \frac{U(y)}{y} \leq \frac{U(x+y)}{x+y} \Longrightarrow xU(x+y)+yU(x+y) \geq (x+y)(U(x)+U(y)).$$

Next, with obvious notation,

$$U_\nu(Z_1\mu_1 + Z_2\mu_2) = \int U(Z_1\rho_1 + Z_2\rho_2)\, d\nu + U'(\infty)\big(Z_1\, \mu_{1,s}[M] + Z_2\, \mu_{2,s}[M]\big);$$

$$U_{Z_1,\nu}(\mu_1) = \frac{1}{Z_1}\int U(Z_1\rho_1)\, d\nu + U'(\infty)\mu_{1,s}[M];$$

$$U_{Z_2,\nu}(\mu_2) = \frac{1}{Z_2}\int U(Z_2\rho_2)\, d\nu + U'(\infty)\mu_{2,s}[M];$$

so part (i) of the lemma follows immediately from (1.7.4). The claim about equality is obvious since it amounts to say that $U(x + y) = U(x) + U(y)$ as soon as either x or y is zero.

The proof of part (ii) is based on a similar type of reasoning. First note that (with the conventions $U(0)/0 = U'(0)$, $U(\infty)/\infty = U'(\infty)$ and μ_s-almost surely, $d\mu/d\nu = +\infty$)

$$U^\beta_{Z_1\pi_1+Z_2\pi_2,\nu}(Z_1\mu_1 + Z_2\mu_2) = \int_{M\times M} U\left(\frac{Z_1\rho_1(x) + Z_2\rho_2(x)}{\beta(x,y)}\right) \frac{\beta(x,y)}{Z_1\rho_1(x) + Z_2\rho_2(x)}\,(Z_1\pi_1 + Z_2\pi_2)(dx\,dy);$$

$$U^{\beta}_{Z_1,\pi_1,\nu}(\mu_1) = \int U\left(\frac{Z_1\rho_1(x)}{\beta(x,y)}\right)\frac{\beta(x,y)}{Z_1\rho_1(x)}\, Z_1\pi_1(dx\,dy);$$

$$U^{\beta}_{Z_2,\pi_2,\nu}(\mu_2) = \int U\left(\frac{Z_2\rho_2(x)}{\beta(x,y)}\right)\frac{\beta(x,y)}{Z_2\rho_2(x)}\, \pi_2(dx\,dy).$$

So the proof of the lemma will be complete if we can show that

$$U\left(\frac{Z_1\rho_1+Z_2\rho_2}{\beta}\right)\frac{\beta}{Z_1\rho_1+Z_2\rho_2}\,(Z_1\pi_1+Z_2\pi_2) \\ \geq U\left(\frac{Z_1\rho_1}{\beta}\right)\frac{\beta}{Z_1\rho_1}\,(Z_1\pi_1) + U\left(\frac{Z_2\rho_2}{\beta}\right)\frac{\beta}{Z_2\rho_2}\,(Z_2\pi_2). \quad (1.7.5)$$

Since $U(r)/r$ is a nondecreasing function of r, if X_1, X_2, p_1, p_2 are any four nonnegative numbers then

$$\frac{U(X_1+X_2)}{X_1+X_2}(p_1+p_2) \geq \frac{U(X_1)}{X_1}\,p_1 + \frac{U(X_2)}{X_2}\,p_2.$$

To recover (1.7.5), it suffices to apply the latter inequality with

$$X_1 = \frac{Z_1\rho_1(x)}{\beta(x,y)}, \quad X_2 = \frac{Z_2\rho_2(x)}{\beta(x,y)},$$

$$p_1 = \frac{d(Z_1\pi_1)}{d(Z_1\pi_1+Z_2\pi_2)}(x,y), \quad p_2 = \frac{d(Z_2\pi_2)}{d(Z_1\pi_1+Z_2\pi_2)}(x,y)$$

and to integrate against $(Z_1\pi_1+Z_2\pi_2)(dx\,dy)$. □

Proof of Theorem 1.7.7. First we observe that U_ν is well-defined on $P_2(M)$ since, if $\mu = \rho\nu + \mu_s$ is the Lebesgue decomposition of a probability measure $\mu \in P(M)$, then

$$U(\rho) \geq -\|U\|_{\mathrm{Lip}} \quad \rho \in L^1(M,\nu).$$

In fact, there is also an upper bound, so U_ν is well-defined on the whole of $P_2(M)$ with values in $\mathbb{R}$. Moreover, by an approximation argument, we may replace the assumptions of weak a.c.c.s. displacement convexity by weak displacement convexity on the whole of $P_2(M)$. (The proof is the same as in [97, Proposition 3.21] (in the compact case) or [133, Theorem 30.5].)

Let μ_0, μ_1 be any two measures in $P_2(M)$, and let Π be an optimal dynamical transference plan between μ_0 and μ_1. Let further

$$\mu_0 = \rho_0\,\nu + \mu_{0,s}$$

be the Lebesgue decomposition of μ_0 with respect to ν. Let $E^{(a)}$ and $E^{(s)}$ be two disjoint Borel subsets of M such that $\rho_0\,\nu$ is concentrated on $E^{(a)}$ and $\mu_{0,s}$ is concentrated on $E^{(s)}$. We decompose Π as

$$\Pi = \Pi^{(a)} + \Pi^{(s)}, \tag{1.7.6}$$

where

$$\Pi^{(a)} := \Pi \llcorner \big\{ \gamma \in \Gamma \mid \gamma(0) \in E^{(a)} \big\}, \qquad \Pi^{(s)} := \Pi \llcorner \big\{ \gamma \in \Gamma \mid \gamma(0) \in E^{(s)} \big\}.$$

Taking the marginals at time t in (1.7.6) we get

$$\mu_t = \mu_t^{(a)} + \mu_t^{(s)}.$$

In the end, we renormalize $\mu_t^{(a)}$ and $\mu_t^{(s)}$ into probability measures: we define

$$Z^{(a)} = \Pi^{(a)}[\Gamma] = \mu_0^{(a)}[M] = \mu_t^{(a)}[M]; \qquad Z^{(s)} = \Pi^{(s)}[\Gamma],$$

and

$$\hat{\Pi}^{(a)} := \frac{\Pi^{(a)}}{Z^{(a)}}, \qquad \hat{\mu}_t^{(a)} := \frac{\mu_t^{(a)}}{Z^{(a)}}; \qquad \hat{\Pi}^{(s)} := \frac{\Pi^{(s)}}{Z^{(s)}}, \qquad \hat{\mu}_t^{(s)} := \frac{\mu_t^{(s)}}{Z^{(s)}}.$$

So

$$\mu_t = Z^{(a)} \hat{\mu}_t^{(a)} + Z^{(s)} \hat{\mu}_t^{(s)}. \tag{1.7.7}$$

We remark that by what we proved in Section 1.5 $\mu_t^{(a)}$ is absolutely continuous for any $t \in [0,1)$, but $\mu_t^{(s)}$ is not necessarily completely singular.

It follows from [133, Theorem 7.29 (v)] that for any $t \in (0,1)$ there is a Borel map F_t such that $F_t(\gamma_t) = \gamma_0$, $\Pi(d\gamma)$-almost surely. Then $\mu_t^{(s)}$ is concentrated on $F_t^{-1}(E^{(s)})$, while $\mu_t^{(a)}$ is concentrated on $F_t^{-1}(E^{(a)})$; so these measures are singular to each other. Then by Lemma 1.7.9 and (1.7.7), for any $t \in (0,1)$,

$$U_\nu(\mu_t) = Z^{(a)} U_{Z^{(a)},\nu}(\hat{\mu}_t^{(a)}) + Z^{(s)} U_{Z^{(s)},\nu}(\hat{\mu}_t^{(s)}). \tag{1.7.8}$$

In the sequel, we focus on part (i) of Theorem 1.7.7, since the reasoning is quite the same for part (ii). By construction and the restriction property of optimal transport [133, Theorem 7.29], $\hat{\Pi}^{(a)}$ is an optimal dynamical transference plan between $\hat{\mu}_0^{(a)}$ and $\hat{\mu}_1^{(a)}$, and the associated Wasserstein geodesic is $(\hat{\mu}_t^{(a)})_{0 \le t \le 1}$. Since by construction $\hat{\mu}_0^{(a)}$ is absolutely continuous, by what we already proved $(\hat{\mu}_t^{(a)})$ is the *unique* Wasserstein geodesic joining $\hat{\mu}_0^{(a)}$ to $\hat{\mu}_1^{(a)}$. Then we can apply the displacement convexity inequality of the functional $U_{Z^{(a)},\nu}$ along that geodesic:

$$U_{Z^{(a)},\nu}(\hat{\mu}_t^{(a)}) \le (1-t)\, U_{Z^{(a)},\nu}(\hat{\mu}_0^{(a)}) + t\, U_{Z^{(a)},\nu}(\hat{\mu}_1^{(a)}) - \frac{\lambda}{2}\, t\,(1-t)\, W_2^2(\hat{\mu}_0^{(a)}, \hat{\mu}_1^{(a)}). \tag{1.7.9}$$

Next, let $\varphi_k \to 0$ be a sequence of positive numbers. From the nonbranching property of $P_2(M)$ [133, Corollary 7.31], there is only one Wasserstein geodesic joining $\hat{\mu}^{(s)}_{\varphi_k}$ to $\hat{\mu}^{(s)}_1$ and it is obtained by reparameterizing $(\hat{\mu}^{(s)}_t)_{\varphi_k \le t \le 1}$ on $[0,1]$ (with an affine reparameterization in t). So we can also apply the displacement convexity inequality of the functional $U_{Z^{(s)},\nu}$ along that geodesic, and get

$$U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_t) \le \left(\frac{1-t}{1-\varphi_k}\right) U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_{\varphi_k}) + \left(\frac{t-\varphi_k}{1-\varphi_k}\right) U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_1) - \frac{\lambda}{2}(t-\varphi_k)(1-t)\, W_2^2(\hat{\mu}^{(s)}_0, \hat{\mu}^{(s)}_1). \quad (1.7.10)$$

(For the latter term we have used the fact that if $(\mu_t)_{0\le t\le 1}$ is any Wasserstein geodesic, then $W_2(\mu_s,\mu_t) = |t-s|\, W_2(\mu_0,\mu_1)$.)

The first term in the right-hand side of (1.7.10) can be trivially bounded by $U'(\infty)$, which coincides with $U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_0)$ since $\hat{\mu}^{(s)}_0$ is totally singular. Indeed, since $\frac{U(r)}{r} \le U'(\infty)$, we have

$$\begin{aligned}
U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_{\varphi_k}) &= \frac{1}{Z^{(s)}} \int_M U\left(Z^{(s)} \hat{\rho}^{(s)}_{\varphi_k}\right) d\nu + U'(\infty)\, \hat{\mu}^{(s)}_{\varphi_k,s}(M) \\
&= \frac{1}{Z^{(s)}} \int_{\{\hat{\rho}^{(s)}_{\varphi_k} > 0\}} U\left(Z^{(s)} \hat{\rho}^{(s)}_{\varphi_k}\right) d\nu + U'(\infty)\, \hat{\mu}^{(s)}_{\varphi_k,s}(M) \\
&= \int_{\{\hat{\rho}^{(s)}_{\varphi_k} > 0\}} \frac{U\big(Z^{(s)} \hat{\rho}^{(s)}_{\varphi_k}\big)}{Z^{(s)} \hat{\rho}^{s}_{\varphi_k}} \hat{\rho}^{(s)}_{\varphi_k}\, d\nu + U'(\infty)\, \hat{\mu}^{(s)}_{\varphi_k,s}(M) \\
&\le \int_{\{\hat{\rho}^{(s)}_{\varphi_k} > 0\}} U'(\infty) \hat{\rho}^{(s)}_{\varphi_k}\, d\nu + U'(\infty)\, \hat{\mu}^{(s)}_{\varphi_k,s}(M) \\
&= U'(\infty)\, \hat{\mu}^{(s)}_{\varphi_k}(M) = U'(\infty).
\end{aligned}$$

Then by passing to the lim inf as $k \to \infty$ in (1.7.10), we recover

$$U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_t) \le (1-t)\, U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_0) + t\, U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_1) - \frac{\lambda}{2} t(1-t)\, W_2^2(\hat{\mu}^{(s)}_0, \hat{\mu}^{(s)}_1). \quad (1.7.11)$$

By combining together (1.7.8), (1.7.9) and (1.7.11), we obtain

$$U_\nu(\mu_t) \le (1-t)\left[Z^{(a)} U_{Z^{(a)},\nu}(\hat{\mu}^{(a)}_0) + Z^{(s)} U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_0)\right] + t\left[Z^{(a)} U_{Z^{(a)},\nu}(\hat{\mu}^{(a)}_1) + Z^{(s)} U_{Z^{(s)},\nu}(\hat{\mu}^{(s)}_1)\right] - \frac{\lambda}{2} t(1-t)\left[Z^{(a)} W_2^2(\hat{\mu}^{(a)}_0, \hat{\mu}^{(a)}_1) + Z^{(s)} W_2^2(\hat{\mu}^{(s)}_0, \hat{\mu}^{(s)}_1)\right]. \quad (1.7.12)$$

The last term inside square brackets can be rewritten as

$$\int d^2(\gamma_0,\gamma_1)\Pi^{(a)}(d\gamma)+\int d^2(\gamma_0,\gamma_1)\Pi^{(s)}(d\gamma)=\int d^2(\gamma_0,\gamma_1)\Pi(d\gamma)=W_2^2(\mu_0,\mu_1).$$

Plugging this back into (1.7.12) and using Lemma 1.7.9, we conclude that

$$U_\nu(\mu_t)\le(1-t)\,U_\nu(\mu_0)+t\,U_\nu(\mu_1)-\frac{\lambda}{2}\,t(1-t)\,W_2^2(\mu_0,\mu_1).$$

This finishes the proof of Theorem 1.7.7. □

Proof of Corollary 1.7.8. Let $U := r \to -r^{1-1/N}$. By the estimates derived in [97, Proposition E.17], U_ν is well-defined on $P_p(M)$. (This is made more explicit in [133, Theorem 17.8 and Example 17.9].)

Let $\mathcal{DC}_n$ be the displacement convex class of order n, that is the class of functions $U\in C^2(0,\infty)\cap C([0,+\infty))$ such that $U(0)=0$ and $\delta^n U(\delta^{-n})$ is a convex function of δ. (See [133, Definition 17.1]). Obviously, $U\in\mathcal{DC}_n$. By [133, Proposition 17.7], there is a sequence $(U^{(\ell)})_{\ell\in\mathbb{N}}$ of Lipschitz functions, all belonging to $\mathcal{DC}_n$, such that $U^{(\ell)}$ converges monotonically to U as $\ell\to\infty$.

Since $U^{(\ell)}$ lies in $\mathcal{DC}_n$, it is by now classical (see [133, Theorem 17.15], which summarizes the works of many authors) that $U_\nu^{(\ell)}$ it is a.c.c.s-displacement convex. By Theorem 1.7.7, this functional is also displacement convex. Then it follows by an easy limiting argument that U_ν itself is displacement convex. □

1.8 A generalization of the existence and uniqueness result

Now we want to generalize this existence and uniqueness result for optimal transport mapping without any integrability assumption on the cost function, adapting the ideas of [107]. We assume that M is an n-dimensional manifold and N a locally compact Polish space. We observe that, without the hypothesis

$$\int_{M\times N} c(x,y)\,d\mu(x)\,d\nu(y)<+\infty,$$

in general the minimization problem

$$C(\mu,\nu):=\inf_{\gamma\in\Pi(\mu,\nu)}\left\{\int_{M\times N} c(x,y)\,d\gamma(x,y)\right\}$$

is ill-posed, as it may happen that $C(\mu,\nu) = +\infty$. Howewer, it is known that the optimality of a transport plan γ is equivalent to the c-cyclical monotonicity of the measure-theoretic support of γ whenever $C(\mu,\nu) < +\infty$ (see [13], [120], [133]), and so one may ask whether the fact that the support of γ is c-cyclically monotone implies that γ is supported on a graph. Moreover one can also ask whether this graph is unique, that is is does not depends on γ, which is the case when the cost is $\mu \otimes \nu$ integrable, as Theorem 1.3.2 tells us. The uniqueness in that case, follows by the fact that the functions φ_n are constructed using a pair of function (φ,ψ) which is optimal for the dual problem, and so they are independent of γ. The result we now want to prove is the following:

Theorem 1.8.1. *Assume that $c : M \times N \to \mathbb{R}$ is lower semicontinuous and bounded from below, and let γ be a plan concentrated on a c-cyclically monotone set. If*

(i) *the family of maps $x \mapsto c(x,y) = c_y(x)$ is locally semi-concave in x locally uniformly in y;*

(ii) *$\frac{\partial c}{\partial x}(x,\cdot)$ is injective on its domain of definition;*

(iii) *and the measure μ gives zero mass to sets with σ-finite $(n-1)$-dimensional Hausdorff measure,*

then γ is concentrated on the graph of a measurable map $T : M \to N$ (existence). Moreover, if $\tilde{\gamma}$ is another plan concentrated on a c-cyclically monotone set, then $\tilde{\gamma}$ is concentrated on the same graph (uniqueness).

Proof. Since the proof of the existence result is the same as in Theorem 1.3.2, we concentrate on the uniqueness part. As we observed before, the difference here with the case of Theorem 1.3.2 is that the function φ_n depends on the pair (φ,ψ), which in this case depends on γ.

Let (φ,ψ) be a pair associated to γ as in Theorem 1.3.2, and let φ_n and B_n be such that γ is concentrated on the graph of the map T determined on B_n by

$$\frac{\partial c}{\partial x}(x,T(x)) = -d_x\varphi_n \quad \text{for } x \in B_n$$

We observe that, thanks to the local compactness of N, in the formula

$$\varphi_n(x) = \sup_{y\in V_n} \psi(y) - c(x,y) \tag{1.8.1}$$

we can assume V_n to be compact. We moreover recall the equality

$$\varphi(x) = \psi(T(x)) - c(x,T(x)) \quad \forall x \in \bigcup_n B_n. \tag{1.8.2}$$

Let now $(\tilde{\varphi}, \tilde{\psi})$ be a pair associated to $\tilde{\gamma}$, and let $\tilde{\varphi}_n$, $\tilde{B}_n$ and $\tilde{T}$ be constructed as above. We need to prove that $T = \tilde{T}$ μ-a.e.
Let us define $C_n := B_n \cap \tilde{B}_n$. Then $\mu(C_n) \nearrow 1$. We want to prove that, if x is a μ-density point of C_n for a certain n, then $T(x) = \tilde{T}(x)$ (we recall that, since $\mu(\cup_n C_n) = 1$, also the union of the μ-density points of C_n is of full μ-measure, see for example [61, Chapter 1.7]).
Let us assume by contradiction that $T(x) \neq \tilde{T}(x)$, that is

$$d_x\varphi_n \neq d_x\tilde{\varphi}_n.$$

Since $x \in \mathrm{supp}(\mu)$, each ball around x must have positive measure under μ. Moreover, the fact that the sets $\{\varphi_n = \varphi\}$ and $\{\tilde{\varphi}_n = \tilde{\varphi}\}$ have μ-density 1 in x implies that the set

$$\{\varphi = \tilde{\varphi}\}$$

has μ-density 0 in x. In fact, as φ_n and $\tilde{\varphi}_n$ are locally semi-convex, up to adding a C^1 function they are concave in a neighborhood of x and their gradients differ at x. So we can apply the non-smooth version of the implicit function theorem proven in [107], which tells us that $\{\varphi_n = \tilde{\varphi}_n\}$ is a set with finite $(n-1)$-dimensional Hausdorff measure in a neighborhood of x (see [107, Theorem 17 and Corollary 19]). So we have

$$\limsup_{r\to 0} \frac{\mu(\{\varphi = \tilde{\varphi}\} \cap B_r(x))}{\mu(B_r(x))} \leq \limsup_{r\to 0} \Bigg[\frac{\mu(\{\varphi \neq \varphi_n\} \cap B_r(x))}{\mu(B_r(x))} + \frac{\mu(\{\varphi_n = \tilde{\varphi}_n\} \cap B_r(x))}{\mu(B_r(x))} + \frac{\mu(\{\tilde{\varphi}_n \neq \tilde{\varphi}\} \cap B_r(x))}{\mu(B_r(x))} \Bigg] = 0.$$

Now, exchanging φ_n with $\tilde{\varphi}_n$ if necessary, we may assume that

$$\mu(\{\varphi_n < \tilde{\varphi}_n\} \cap B_r(x)) \geq \frac{1}{3}\mu(B_r(x)) \quad \text{for } r > 0 \text{ sufficiently small,}$$

which implies

$$\mu(\{\varphi < \tilde{\varphi}\} \cap B_r(x)) \geq \frac{1}{4}\mu(B_r(x)) \quad \text{for } r > 0 \text{ sufficiently small.} \tag{1.8.3}$$

Let us define $A := \{\varphi < \tilde{\varphi}\}$, $A_n := \{\varphi_n < \tilde{\varphi}_n\}$, $E_n := A \cap A_n \cap C_n$. Since the sets $\{\varphi_n = \varphi\}$ and $\{\tilde{\varphi}_n = \tilde{\varphi}\}$ have μ-density 1 in x, and x is a μ-density point of C_n, we have

$$\lim_{r\to 0} \frac{\mu((A \setminus E_n) \cap B_r(x))}{\mu(B_r(x))} = 0,$$

and so, by (1.8.3), we get

$$\mu(E_n \cap B_r(x)) \geq \frac{1}{5}\mu(B_r(x)) \quad \text{for } r > 0 \text{ sufficiently small.} \tag{1.8.4}$$

Now, arguing as in the proof of the Aleksandrov's lemma (see [107, Lemma 13]), we can prove that

$$X := \tilde{T}^{-1}(T(A)) \subset A$$

and $X \cap E_n$ lies a positive distance from x. In fact let us assume, without loss of generality, that

$$\varphi(x) = \varphi_n(x) = \tilde{\varphi}(x) = \tilde{\varphi}_n(x) = 0, \quad d_x\varphi_n \neq d_x\tilde{\varphi}_n = 0.$$

To obtain the inclusion $X \subset A$, let $z \in X$ and $y := \tilde{T}(z)$. Then $y = T(m)$ for a certain $m \in A$. For any $w \in M$, recalling (1.8.2), we have

$$\varphi(w) \leq c(m, y) - c(w, y) + \varphi(m),$$

$$\tilde{\varphi}(m) \leq c(z, y) - c(m, y) + \tilde{\varphi}(z).$$

Since $\varphi(m) < \tilde{\varphi}(m)$ we get

$$\varphi(w) < c(z, \tilde{T}(z)) - c(w, \tilde{T}(z)) + \tilde{\varphi}(z) \quad \forall w \in M.$$

In particular, taking $w = z$, we obtain $z \in A$, that proves the inclusion $X \subset A$.
Let us suppose now, by contradiction, that there exists a sequence $(z_k) \subset X \cap E_n$ such that $z_k \to x$. Again there exists m_k such that $\tilde{T}(z_k) = T(m_k)$. As $d_x\tilde{\varphi}_n = 0$, the closure of the superdifferential of a semi-concave function implies that $d_{z_k}\tilde{\varphi}_n \to 0$. We now observe that, arguing exactly as above with φ_n and $\tilde{\varphi}_n$ instead of φ and $\tilde{\varphi}$, using (1.8.1), (1.8.2), and the fact that $\varphi = \varphi_n$ and $\tilde{\varphi} = \tilde{\varphi}_n$ on C_n, one obtains

$$\varphi_n(w) < c(z_k, \tilde{T}(z_k)) - c(w, \tilde{T}(z_k)) + \tilde{\varphi}_n(z_k) \quad \forall w \in M.$$

Taking w sufficiently near to x, we can assume that we are in $\mathbb{R}^n \times N$. We now remark that, since $z_k \in E_n \subset \tilde{D}_n$, $\tilde{T}(z_k)$ vary in a compact subset of N. So, by hypothesis *(i)* on c, we can find a common modulus of continuity ω in a neighborhood of x for the family of uniformly semi-concave functions $z \mapsto c(z, \tilde{T}(z_k))$. Then, we get

$$\begin{aligned}\varphi_n(w) &< -\frac{\partial c}{\partial x}(z_k, \tilde{T}(z_k))(w - z_k) + \omega(|w - z_k|)|w - z_k| + \tilde{\varphi}_n(z_k) \\ &= d_{z_k}\tilde{\varphi}_n(w - z_k) + \omega(|w - z_k|)|w - z_k| + \tilde{\varphi}_n(z_k).\end{aligned}$$

Letting $k \to \infty$ and recalling that $d_{z_k}\tilde{\varphi}_n \to 0$ and $\tilde{\varphi}_n(x) = \varphi_n(x) = 0$, we obtain

$$\varphi_n(w) - \varphi_n(x) \le \omega(|w - x|)|w - x| \Rightarrow d_x\varphi_n = 0,$$

which is absurd.
Thus there exists $r > 0$ such that $B_r(x) \cap E_n$ and $X \cap E_n$ are disjoint, and (1.8.4) holds. Defining now $Y := T(A)$, by (1.8.4) we obtain

$$\begin{aligned}\nu(Y) = \mu(T^{-1}(Y)) \ge \mu(A) = \mu(E_n) + \mu(A \setminus E_n) \ge \mu(B_r(x) \cap E_n)\\ + \mu(X \cap E_n) + \mu(X \setminus E_n) = \mu(B_r(x) \cap E_n) + \mu(X) \ge \frac{1}{5}\mu(B_r(x)) + \nu(Y),\end{aligned}$$

which is absurd. □

Let now consider the special case $N = M$, with M a complete manifold. As shown in Paragraph 1.4, the above theorem applies in the following cases:

1. $c : M \times M \to \mathbb{R}$ is defined by

$$c(x, y) := \inf_{\gamma(0)=x,\ \gamma(1)=y} \int_0^1 L(\gamma(t), \dot{\gamma}(t))\, dt,$$

   where the infimum is taken over all the continuous piecewise C^1 curves, and the Lagrangian $L(x, v) \in C^2(TM, \mathbb{R})$ is C^2-strictly convex and uniform superlinear in v, and verifies an uniform boundeness in the fibers;

2. $c(x, y) = d^p(x, y)$ for any $p \in (1, +\infty)$, where $d(x, y)$ denotes a complete Riemannian distance on M.

Chapter 2

The irrigation problem

2.1 Introduction

[1] The variety of structures arising in nature is extraordinary. By exploring the relationship between form and function, D'Arcy Thompson, in his pioneering work [53], tries to find common principles behind the varied phenomena (physical, chemical, biological, short or long time scale, etc.) that interact to give birth to these structures. Indeed, despite the complexity of nature, the approach of retaining only a small but decisive set of parameters and principles to model the phenomenon at the origin of a given structure can be successful. See for example [113] or consider the work of Turing on morphogenesis that led him to explain the appearance of heterogeneous spatial patterns in terms of reaction-diffusion mechanisms [131].

Recently, such an approach was taken to model branched networks that achieve a transport from a source to a target. Such networks are everywhere in nature (plants and trees, river basins, bronchial and cardiovascular systems) and in man designed structures (communication networks, electric power supply, water distribution or drainage networks). The common function of such networks is to transport some goods from an initial distribution (the supply) to another (the demand). Following D'Arcy Thompson, it is desirable to tie a link between this unity of form (branched networks) and this unity of function (transporting goods from a supply to a demand). This was done in [82, 98, 135, 25, 24, 29] by considering cost functions that encode the efficiency of a transport induced by some structure. Branched structures, as the one observed in nature, then arise as the optimal structures along which the transport takes place.

A simple but crucial principle was incorporated in the design of all the cost functions

[1] This chapter is based on a joint work with Marc Bernot [27].

used by these authors. This principle states that it is more efficient to transport mass in a grouped way rather than in a separate way. To embed this principle, the previously mentioned costs incorporate a parameter $\alpha \in [0,1]$ and make use of the concavity of $x \mapsto x^\alpha$. The idea is that for positive masses m_1 and m_2, we have $(m_1+m_2)^\alpha \leq m_1^\alpha + m_2^\alpha$, so that the particles are interested in moving together in order to lower the cost (see for example the role of α in (2.1.1)). This effect gets stronger as α decreases, while the limit case $\alpha = 1$ gives no importance to the grouping of particles.

We now briefly review the different costs and descriptions of branched structures that have been introduced so far. We then introduce a new dynamical cost functional, and enlight the advantages it has over other models.

The model described by Gilbert in [82] consists in finite directed weighed graph G with straight edges $E(G)$ and a weight function $w : E(G) \to (0, \infty)$. The graph G connects sources $\mu^+ = \sum_{i=1}^k a_i \delta_{x_i}$ and targets $\mu^- = \sum_{j=1}^l b_j \delta_{y_j}$ with $\sum_i a_i = \sum_j b_j$, $a_i, b_j \geq 0$, and is required to satisfy Kirchhoff's law at each vertex. The cost of G is defined to be:

$$M^\alpha(G) = \sum_{e \in E(G)} w(e)^\alpha \mathcal{H}^1(e). \tag{2.1.1}$$

In [135], Xia extends this model to a continuous framework using Radon vector measures. In both these models, the objects and their costs are static in the sense that no "particle" is actually transported along the structure, and the cost depends only on the geometry of the network.

In [98, 25, 24], a different kind of object, called traffic plan and denoted by χ, is considered. In this framework, all particles are indexed by the set $\Omega := [0,1]$, and to each $\omega \in \Omega$ is associated a 1-Lipschitz path $\chi(\omega, \cdot)$ in $\mathbb{R}^N$. This is a Lagrangian description of the dynamic of particles that can be encoded by the image measure $\mathbf{P}_\chi$ of the map $\omega \mapsto \chi(\omega, \cdot)$ (which is therefore a measure on the set of 1-Lipschitz paths). This measure induces a network structure similar to the one considered by Xia. To each traffic plan is associated a cost E^α which depends only on its network structure (see Definition 2.2.4) and, whenever it is finite, is the same as the one considered by Xia. Thus, though a traffic plan is a dynamical object, its cost is static.

In [29], Brancolini, Buttazzo and Santambrogio consider an Eulerian formulation of the problem, describing a transport from μ^+ to μ^- as a path in the space of measures. The cost of such a path is defined as the length induced by a degenerate Riemannian metric in the space of probability measures. More precisely, the cost of a path $\mu(t)$ is given by

$$\int_0^1 J(\mu(t)) |\mu'|(t) \, dt,$$

where J is a functional in the space of probability measures and $|\mu'|$ denotes the metric derivative (for the Wasserstein distance) of the path. Both the object and the cost are dynamical in this model.

All the above described models propose structures that transport a measure μ^+ to a measure μ^- and associate a cost to this structure. This leads to consider what is called *the irrigation problem* by some authors [25, 24, 26], i.e., given two measures μ^+ and μ^-, the problem of minimizing the cost among structures transporting μ^+ to μ^-. In the case of the traffic plan model, an additional problem can be considered, namely the *who goes where problem* [25, 24]. The latter problem consists in looking for an optimal structure that achieves a given transference plan. In other words, rather than only prescribing the initial an final distribution of masses as in the irrigation problem, one also prescribes the coupling between initial and final positions of each particle. As an example, one can think about the case where the initial distribution represents the habitations, and the final distribution stands for the workplaces. In this case, it is natural to constrain each inhabitant to go from his habitation to his workplace, and so the problem is to find the best itinerary he can follow.

Here, we consider the Lagrangian formulation given in [98, 25]. This choice is motivated by the fact that traffic plans permit to recover other descriptions. Indeed, given a traffic plan χ, one can always canonically define both a structure similar to the one of Xia, and a path in the space of measures by considering the time marginals of its induced measure $\mathbf{P}_\chi$. We consider general costs of the form

$$\mathcal{C}(\chi) := \int_\Omega \int_{\mathbb{R}^+} c(\chi,\omega,t)|\dot{\chi}(\omega,t)|\, dt\, d\omega. \tag{2.1.2}$$

The advantage of the Lagrangian formulation with respect to the Eulerian one is to allow to define costs of the above form in which one can take care of the speed of each single particle, so that only moving particles contribute to the total cost.

What we propose, is to give a cost to the actual "dynamical" transport of mass from μ^+ to μ^- that is induced by χ. To obtain such a cost, it is natural to require $c(\chi,\omega,t)$ to be local in space-time. By this property, we mean that $c(\chi,\omega,t)$ only takes into account the particles that are located at the point $\chi(\omega,t)$ at time t. In [25] is considered a cost $c(\chi,\omega,t)$ depending on the total mass of particles passing through the point $\chi(\omega,t)$ at some time (see Definition 2.2.4). Since it takes into account only the global trajectories of particles but not their local dynamics, this cost is local in space but not in time. The associated functional E^α thus quantifies the cost of the structure achieving the transport, rather than the cost of the transport itself. In other words, we could also say that E^α evaluates the cost of permanent regime connecting μ^+ to μ^-, rather than the cost of a dynamical transport from μ^+ to μ^-. The elementary cost c we introduce in Definition 2.3.3 has the desired locality property, and we denote by C^α the induced cost via formula (2.1.2). It is possible to extend the time domain by replacing $\mathbb{R}^+$ with $\mathbb{R}$ in (2.1.2), and we denote by $E^\alpha_{\mathbb{R}}$ and $C^\alpha_{\mathbb{R}}$ the costs corresponding to E^α and C^α.

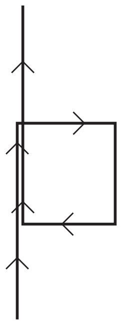

Figure 2.1: In the case of the static cost E^α in [25], a portion of a path where it overlaps with itself contributes only once to the total cost, whereas the locality in time of the model we propose gives the expected cost.

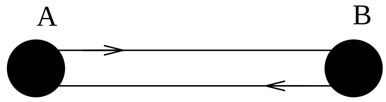

Figure 2.2: The best way to switch two equal masses between two points A and B is to transport the mass at A to position B and the mass at B to position A along the segment joining them. For such a structure, the C^α cost we propose distinguishes between trajectories going from A to B and from B to A, which is not the case of the E^α cost. Thus, the C^α cost is more realistic for the "who goes where" problem.

We illustrate the advantage of such a "dynamical" cost with respect to the static one in [25] on two examples:

- It gives a more realistic cost to an overlapping path. Indeed, in the case of the static cost in [25], a path that follows the same circuit twice contributes to the cost once, while the locality in time of the model we propose gives the expected cost (see figure 2.1).

- It is more appropriate for the "who goes where problem". Let us consider the problem of two equal masses m located at points A and B, which represent both the source and the target distribution, and where the transference plan constraint consists in switching the two masses. In this case, the solution to this "who goes where" problem is to transport the mass in A to position B and the mass in B to position A along the segment joining them. For such a structure, the E^α cost does not distinguish between trajectories going from A to B and from B to A. Indeed, the E^α cost of this structure is $|A-B|(2m)^\alpha$, while the natural one would be $2|A-B|m^\alpha$. This is exactly the cost given by C^α (see figure 2.2).

We will consider the irrigation problems for all the just mentioned costs. As it will be proved in Section 2.5, the two irrigation problems with costs E^α and C^α are equivalent if μ^+ is a finite atomic, while the equivalence for $E^\alpha_\mathbb{R}$ and $C^\alpha_\mathbb{R}$ always holds. More precisely, in these cases, we will prove that any minimizer for the dynamical cost is an E^α-minimizer, and that moreover, up to reparameterization, the converse is true (see the remarks after Theorem 2.5.2). Since the cost $E^\alpha(\chi)$ is invariant by reparameterization of the traffic plan χ, while $C^\alpha(\chi)$ in general is not, this fact will tell us in particular that the cost C^α has the feature to select, among all the possible reparameterization of an optimal traffic plan χ, some particular ones, in which particles actually move in a grouped way.

Given two measures μ^+ and μ^-, let us define

$$E^\alpha(\mu^+, \mu^-) := \inf E^\alpha(\chi),$$

where the infimum is taken over all traffic plans transporting μ^+ onto μ^- (the same can be done with C^α, $E^\alpha_{\mathbb{R}}$ and $C^\alpha_{\mathbb{R}}$). By the above formula, one obtains a one-parameter family of distances between measures, each of them inducing the weak-$*$ topology. It turns out that the continuity of the function $\alpha \mapsto E^\alpha(\mu^+, \mu^-)$ is related to the following stability property: given a converging sequence of traffic plans χ_n, respectively optimal for the value α_n, its limit is optimal for the limit value of α_n. In particular, considering a sequence $\alpha_n \to 1$, one would obtain the convergence of optimal structures to an optimal structure for the 1-Wasserstein distance. It is therefore of interest to study the α dependence of $E^\alpha(\mu^+, \mu^-)$. This α dependence will be shown in Section 2.6 to be continuous if $\alpha \in [1 - \frac{1}{N}, 1]$ (N being the dimension of the ambient space).

The plan is as follows. In Section 2.2, we recall the main definitions and results concerning traffic plans. In Section 2.3, we consider the energy functional of [25] in a more general framework for which we obtain a general lower semicontinuity result. Then we define a new dynamical (in the sense previously discussed) cost functional and obtain a partial result of existence of a "dynamical" optimal traffic plan for the irrigation problem. We can however obtain a more complete existence result by studying the properties of E^α-minimizers. Indeed, in Section 2.4, we prove that any E^α-optimal traffic plan can be suitably reparameterized. From this fact, we deduce in Section 2.5 that the cost of optimal traffic plans and dynamical optimal traffic plans are the same, and that any E^α-optimal traffic plan can be reparameterized so that it is becomes optimal also for the dynamical cost C^α (this is always true for $E^\alpha_{\mathbb{R}}$ and $C^\alpha_{\mathbb{R}}$, while for E^α and C^α we need μ^+ to be finite atomic). Finally, in Section 2.6, we prove continuity results of $E^\alpha(\mu^+, \mu^-)$ with respect to α, for fixed μ^+ and μ^-. As we already said above, this implies that limits of optimal (for different values of α) traffic plans are still optimal for the limit value.

2.2 Traffic plans

In this section, we recall main definitions and results concerning traffic plans (see [98, 135, 24, 25, 26]). Let X be some compact convex N-dimensional set in $\mathbb{R}^N$. We shall denote by $\mathscr{L}^1(A)$ the Lebesgue measure of a measurable set $A \subset \mathbb{R}$, and by $\mathrm{Lip}_1(\mathbb{R}^+, X)$ the space of 1-Lipschitz curves in X with the metric of uniform convergence on compact sets of $\mathbb{R}^+$.

Definition 2.2.1. Let $\Omega = [0,1]$. A traffic plan is a measurable map $\chi : \Omega \times \mathbb{R}^+ \to X$ such that for all ω, $t \mapsto \chi(\omega, t)$ is 1-Lipschitz, and constant for t sufficiently large. Without risk of ambiguity, we shall call *fiber* both the path $\chi(\omega, \cdot)$ and $\omega \in \Omega$. We denote by $\mathbf{P}_\chi$ the law of $\omega \mapsto \chi(\omega) \in \mathrm{Lip}_1(\mathbb{R}^+, X)$ defined by $\mathbf{P}_\chi(E) := \mathscr{L}^1(\chi^{-1}(E))$ for every Borel set $E \subset \mathrm{Lip}_1(\mathbb{R}^+, X)$.

We remark that in the sequel we will also need to consider the restriction of a traffic plan to a certain subset of fibers $\Omega' \subset \Omega$. By abuse of notation, though Ω' will not be of unit mass, we will still call $\chi \llcorner \tilde{\Omega}'$ a traffic plan.

Definition 2.2.2. Two traffic plans χ and χ' are said to be equivalent if $\mathbf{P}_\chi = \mathbf{P}_{\chi'}$. In all the following a "traffic plan" means as well the equivalence class of some χ. All proven properties of a traffic plan will be true for any representative up to the addition or removal of a set of fibers with zero measure.

Stopping time, irrigated measures, transference plan

If $\chi : \Omega \times \mathbb{R}^+ \to X$ is a traffic plan, define its stopping time by

$$T_\chi(\omega) := \inf\{t \geq 0 : \chi(\omega) \text{ is constant on } [t, \infty)\}.$$

Let us denote the initial and final point of a fiber ω by $\tau(\omega) = \chi(\omega, 0)$ and $\sigma(\omega) = \chi(\omega, T_\chi(\omega))$. To any χ, one can associate its irrigating and irrigated measure respectively defined by

$$\mu^+(\chi)(A) := \tau_\# \mathbf{P}_\chi(A) = \mathscr{L}^1(\{\omega : \chi(\omega, 0) \in A\}),$$
$$\mu^-(\chi)(A) := \sigma_\# \mathbf{P}_\chi(A) = \mathscr{L}^1(\{\omega : \chi(\omega, T_\chi(\omega)) \in A\}),$$

where A is any Borel subset of $\mathbb{R}^N$.

Energy of a traffic plan

Definition 2.2.3. Let $\chi : \Omega \times \mathbb{R}^+ \to X$ be a traffic plan. Define the *path class of* $x \in \mathbb{R}^N$ *in* χ as the set

$$\Omega_x^\chi := \{\omega : x \in \chi(\omega, \mathbb{R})\},$$

and the *multiplicity of* χ *at* x by $\theta_\chi(x) = \mathscr{L}^1(\Omega_x^\chi)$. For simplicity, we shall write $\Omega_x := \Omega_x^\chi$, whenever the underlying traffic plan χ is not ambiguous.

We use the convention that $0^{\alpha-1} = +\infty$ for $\alpha \in [0,1)$.

Definition 2.2.4. Let $\alpha \in [0,1]$. We call *energy* of a traffic plan $\chi : \Omega \times \mathbb{R}^+ \to X$ the functional

$$E^\alpha(\chi) = \int_\Omega \int_{\mathbb{R}^+} \theta_\chi(\chi(\omega, t))^{\alpha-1} |\dot{\chi}(\omega, t)| \, dt \, d\omega. \tag{2.2.1}$$

Let μ^+, μ^- be two probability measures in X. Denote by $\mathrm{TP}(\mu^+, \mu^-)$ the set of traffic plans χ such that $\mu^+(\chi) = \mu^+$ and $\mu^-(\chi) = \mu^-$. If $C > 0$, call TP_C the set of traffic plans such that $\int_\Omega T_\chi(\omega)\, d\omega \le C$ and $\mathrm{TP}_C(\mu^+, \mu^-) := \mathrm{TP}(\mu^+, \mu^-) \cap \mathrm{TP}_C$.

Convergence

Definition 2.2.5. We say that a sequence of traffic plans χ_n *converges* to a traffic plan χ if $\mathbf{P}_{\chi_\mathbf{n}}$ weakly-$*$ converges to $\mathbf{P}_\chi$, or equivalently if the random variables χ_n converge in law to χ.

Definition 2.2.6. We say that a sequence of traffic plans χ_n *fiber converges* to a traffic plan χ if $\chi_n(\omega)$ converges to $\chi(\omega)$ uniformly on compact subsets of $\mathbb{R}^+$ for every $\omega \in \Omega$ (this is stronger than the usual almost sure convergence of random variables).

Remark 2.2.7. *By Skorokhod theorem (see Theorem 11.7.2 [57]) χ_n converges to χ if and only if there exist $\tilde{\chi}_n$ and $\tilde{\chi}$ equivalent to χ_n and χ respectively and such that $\tilde{\chi}_n(\omega)$ fiber converges to $\tilde{\chi}(\omega)$.*

Proposition 2.2.8. *Up to a subsequence, any sequence of traffic plans χ_n in TP_C converges to a traffic plan χ. In addition, $\mu^+(\chi_n) \rightharpoonup \mu^+(\chi)$ and $\mu^-(\chi_n) \rightharpoonup \mu^-(\chi)$.*

Existence of minimizers

The optimization problem we are interested in is the *irrigation problem*, i.e. the problem of minimizing $E^\alpha(\chi)$ in $\mathrm{TP}(\mu^+, \mu^-)$. The following results are proved in [24, 98, 25].

Theorem 2.2.9. *If $C > 0$ and $\chi_n : \Omega \times \mathbb{R}^+ \to X$ is a sequence in TP_C converging to the traffic plan χ, then*

$$E^\alpha(\chi) \le \liminf_n E^\alpha(\chi_n).$$

We notice that the cost $E^\alpha(\chi)$ is invariant by time-reparameterization of χ. Therefore, one can always reparameterize χ so that $|\dot{\chi}(\omega, t)| = 1$ for all $t \in (0, T_\chi(\omega))$ without changing the cost. In this case, since $\theta_\chi^{\alpha-1} \ge 1$, one gets $\int_\Omega T_\chi(\omega) d\omega \le E^\alpha(\chi)$. Thus, if χ_n is a sequence of traffic plan with a uniformly bounded E^α cost, it is in TP_C up to reparameterization for C big enough. By Proposition 2.2.8 and Theorem 2.2.9, the direct method of the calculus of variations ensures the existence of an optimal traffic plan in $\mathrm{TP}(\mu^+, \mu^-)$.

Corollary 2.2.10. *The problem of minimizing $E^\alpha(\chi)$ in $\mathrm{TP}(\mu^+, \mu^-)$ admits a solution.*

Definition 2.2.11. A traffic plan χ is said to be *optimal for the irrigation problem* if it is of minimal cost in $\mathrm{TP}(\mu^+(\chi), \mu^-(\chi))$.

Let

$$E^\alpha(\mu^+,\mu^-) := \min_{\mathrm{TP}(\mu^+,\mu^-)} E^\alpha(\chi).$$

As proved in [25], there is an optimal traffic plan in $\mathrm{TP}(\mu^+,\mu^-)$ which is *loop-free*, i.e. for almost any $\omega \in \Omega$, the map $\chi(\omega,\cdot)$ is one to one in $[0, T_\chi(\omega)]$. Moreover, using Propositions 6.4 and 6.6 in [25], given any optimal traffic plan with finite energy there is an equivalent loop-free traffic plan with the same energy, hence optimal. Thus, without loss of generality, we may assume that optimal traffic plans are loop-free.

The triangle inequality for the cost E^α holds (just think of concatenating traffic plans [26]):

Proposition 2.2.12. *Let μ_0, μ_1 and μ_2 be probability measures. We have the triangle inequality*

$$E^\alpha(\mu_0,\mu_2) \le E^\alpha(\mu_0,\mu_1) + E^\alpha(\mu_1,\mu_2).$$

Stability with respect to μ^+ and μ^-

The following results were first proved in a slightly different framework by Xia [135], and their proofs adapt immediately to traffic plans (see [24]). We remark that, here and in the sequel of the chapter, by atomic measure we mean a finite sum of delta measures.

Let C be a cube with edge length L and center c. Let ν be a probability measure on the compact set X where $X \subset C$. We may approximate ν by atomic measures as follow. For each i, let

$$\mathcal{C}_i := \{C_i^h : h \in \mathbb{Z}^N \cap [0,2^i)^N\}$$

be a partition of C into cubes of edge length $\frac{L}{2^i}$. Now, for each $h \in \mathbb{Z}^N \cap [0,2^i)^N$, let c_i^h be the center of C_i^h and $m_i^h = \nu(C_i^h)$ be the ν mass of the cube C_i^h.

Definition 2.2.13. We define the dyadic approximation of ν as

$$A_i(\nu) := \sum_{h\in\mathbb{Z}^N\cap[0,2^i)^N} m_i^h \delta_{c_i^h}.$$

We observe that the measures $A_i(\nu)$ weakly-$*$ converge to ν.

Proposition 2.2.14. *Let $\alpha \in (1-\frac{1}{N},1]$. Let ν be a probability measure with support in a cube centered at c and of edge length L. We have*

$$E^\alpha(A_n(\nu),\nu) \le \frac{2^{n(N(1-\alpha)-1)}}{2^{1-N(1-\alpha)}-1}\frac{\sqrt{N}L}{2}.$$

In particular, $E^\alpha(A_n(\nu),\nu) \to 0$ locally uniformly in α for all ν when $n \to \infty$.

By this result and Theorem 2.2.9, it is not difficult to prove that the cost E^α metrizes the weak-$*$ convergence for $\alpha \in (1 - \frac{1}{N}, 1]$.

Lemma 2.2.15. *Let $\alpha \in (1 - \frac{1}{N}, 1]$. A sequence of probability measures ν_n weakly-$*$ converges to ν if and only if $E^\alpha(\nu_n, \nu) \to 0$ when $n \to \infty$.*

Corollary 2.2.16. *Let $\alpha \in (1 - \frac{1}{N}, 1]$. If χ_n is a sequence of optimal traffic plans for the irrigation problem and $\chi_n \to \chi$, then χ is optimal.*

Moreover, by Proposition 2.2.14, $E^\alpha(\mu^+, \mu^-)$ is always finite for $\alpha \in (1 - \frac{1}{N}, 1]$.

Regularity

The following regularity results were proved in [26].

Proposition 2.2.17. *Let μ^+ and μ^- be atomic probability measures and $\alpha \in [0, 1]$. An optimum for the irrigation problem is a finite tree made of segments (in the sense that the fibers $\chi(\omega, \cdot)$, once parameterized by arc lengths, describe a finite set of piecewise linear curves).*

Theorem 2.2.18. *Let $\alpha \in (1 - \frac{1}{N}, 1)$ and let χ be an optimal traffic plan in $\mathrm{TP}(\mu^+, \mu^-)$. Assume that the supports of μ^+ and μ^- are at positive distance. In any closed ball $B(x, r)$ not meeting the supports of μ^+ and μ^-, the traffic plan has the structure of a finite graph.*

Extension of the time domain

In Sections 2.4 and 2.5, we will consider traffic plans defined on $\Omega \times \mathbb{R}$. All the notions introduced above are easy to generalize, and we shall denote by $\mathrm{TP}_{\mathbb{R}}(\mu^+, \mu^-)$ the set of extended traffic plans from μ^+ and μ^- and $E^\alpha_{\mathbb{R}}$ the corresponding cost. We denote by $\mathrm{TP}_{\mathbb{R},C}(\mu^+, \mu^-)$ the traffic plans $\chi \in \mathrm{TP}_{\mathbb{R}}(\mu^+, \mu^-)$ such that $\int_\Omega T_\chi(\omega) d\omega \leq C$, where for a traffic plan in $\mathrm{TP}_{\mathbb{R}}$

$$T_\chi(\omega) := \inf\{t + s : t, s \geq 0,\ \chi(\omega) \text{ is constant on } (-\infty, -s] \cup [t, \infty)\}.$$

Any traffic plan $\chi \in \mathrm{TP}_{\mathbb{R}}$ can be shifted in time so that it can be seen as a traffic plan in TP and the corresponding $E^\alpha_{\mathbb{R}}$ and E^α costs are the same. Thus, from the point of view of the irrigation problem, the two formalisms yield the same optimal objects. However, the introduction of this extended model is made necessary for the study of the dynamical framework we propose, since the dynamical cost we will consider is not invariant by time-reparameterization.

2.3 Dynamic cost of a traffic plan

Let χ be a traffic plan and $c(\chi,\omega,t)$ the elementary cost due to the particle ω along the fiber $\chi(\omega)$ at time t. We define a general cost function $\mathcal{C}$ of a traffic plan χ as follows:

$$\mathcal{C}(\chi) := \int_\Omega \int_{\mathbb{R}^+} c(\chi,\omega,t)|\dot{\chi}(\omega,t)|\,dt\,d\omega. \tag{2.3.1}$$

The choice $c(\chi,\omega,t) = \theta_\chi(\chi(\omega,t))^{\alpha-1}$ yields the energy of a traffic plan given by Definition 2.2.4. In this section, we first prove that for a large class of elementary costs $c(\chi,\omega,t)$, the cost of a traffic plan $\mathcal{C}(\chi)$ is lower semicontinuous. Then, we introduce a dynamical elementary cost (see the introduction for the meaning of dynamical) for which the corresponding cost $\mathcal{C}$ is lower semicontinuous. This yields the existence of a minimizer for the dynamical irrigation problem.

Proposition 2.3.1. *Let* $c : \mathrm{TP}(\mu^+,\mu^-)\times\Omega\times\mathbb{R}^+ \to \mathbb{R}^+$ *such that* $c(\cdot,\omega,\cdot)$ *is lower semicontinuous (with respect to the fiber convergence on traffic plans and the usual topology in* $\mathbb{R}^+$*) for all* ω*. If* $\chi_n : \Omega\times\mathbb{R}^+ \to X$ *fiber converges to the traffic plan* χ*, then*

$$\mathcal{C}(\chi) \le \liminf_n \mathcal{C}(\chi_n).$$

Proof. Let us set $c^\lambda(\chi,\omega,t) := \inf_{s\ge 0}\{c(\chi,\omega,s)+\lambda|t-s|\}$. Since $c(\chi,\omega,\cdot)$ is lower semicontinuous, it is classical (see [10]) that $c^\lambda(\chi,\omega,\cdot)$ is λ-Lipschitz and that

$$c(\chi,\omega,t) = \sup_\lambda c^\lambda(\chi,\omega,t).$$

Let us prove that $c^\lambda(\cdot,\omega,t)$ is lower semicontinuous for all ω and t. Let $\chi_n \to \chi$, and, for fixed ω and t, assume that up to a subsequence the liminf of $c^\lambda(\chi_n,\omega,t)$ is indeed a limit. Now, for each n, take t_n such that

$$c^\lambda(\chi_n,\omega,t) \ge c(\chi_n,\omega,t_n) + \lambda|t-t_n| - \frac{1}{n}.$$

If $t_n \to +\infty$, since c is non-negative,

$$\lim_n c^\lambda(\chi_n,\omega,t) \ge \liminf_n \lambda|t-t_n| = +\infty \ge c^\lambda(\chi,\omega,t).$$

Otherwise, up to a subsequence, we can assume $t_n \to t_\infty$, so that $\liminf_n c(\chi_n,\omega,t_n) \ge c(\chi,\omega,t_\infty)$. Therefore,

$$\lim_n c^\lambda(\chi_n,\omega,t) \ge \liminf_n c(\chi_n,\omega,t_n) + \lambda|t-t_n| \ge c(\chi,\omega,t_\infty) + \lambda|t-t_\infty| \ge c^\lambda(\chi,\omega,t).$$

Let us fix now $T > 0$ and $\varepsilon > 0$, and let us consider $0 = t_1 \le \ldots \le t_i \le \ldots \le t_k = T$ such that $|t_{i+1} - t_i| \le \varepsilon$. Since $c^\lambda(\chi, \omega, \cdot)$ is λ-Lipschitz, $|\chi(\omega, \cdot)|$ is 1-Lipschitz, and $\chi \mapsto \int |\dot\chi(\omega, t)| dt$ and $\chi \mapsto c^\lambda(\chi, \omega, t)$ are lower semicontinuous for the fiber convergence, we have:

$$\liminf_n \int_{[0,T]} c^\lambda(\chi_n, \omega, t)|\dot\chi_n(\omega, t)|\, dt \ge \sum_i \left[\liminf_n c^\lambda(\chi_n, \omega, t_i) \int_{t_i}^{t_{i+1}} |\dot\chi_n(\omega, t)|\, dt - \lambda\varepsilon(t_{i+1} - t_i)\right]$$
$$\ge \sum_i \left[c^\lambda(\chi, \omega, t_i) \int_{t_i}^{t_{i+1}} |\dot\chi(\omega, t)|\, dt - \lambda\varepsilon(t_{i+1} - t_i)\right] \ge \int_{[0,T]} c^\lambda(\chi, \omega, t)|\dot\chi(\omega, t)|\, dt - 2\lambda\varepsilon T.$$

This being true for all ε, we get for a.e. ω and all $T > 0$,

$$\liminf_n \int_{\mathbb{R}^+} c(\chi_n, \omega, t)|\dot\chi_n(\omega, t)|\, dt \ge \liminf_n \int_{[0,T]} c(\chi_n, \omega, t)|\dot\chi_n(\omega, t)|\, dt$$
$$\ge \liminf_n \int_{[0,T]} c^\lambda(\chi_n, \omega, t)|\dot\chi_n(\omega, t)|\, dt \ge \int_{[0,T]} c^\lambda(\chi, \omega, t)|\dot\chi(\omega, t)|\, dt.$$

Then, by Fatou's lemma,

$$\liminf_n \mathcal{C}(\chi_n) = \liminf_n \int_\Omega \int_{\mathbb{R}^+} c(\chi_n, \omega, t)|\dot\chi_n(\omega, t)|\, dt\, d\omega$$
$$\ge \int_\Omega \liminf_n \int_{\mathbb{R}^+} c(\chi_n, \omega, t)|\dot\chi_n(\omega, t)|\, dt\, d\omega \ge \int_\Omega \int_{[0,T]} c^\lambda(\chi, \omega, t)|\dot\chi(\omega, t)|\, dt\, d\omega,$$

and we conclude thanks to the monotone convergence theorem. □

We now define the dynamical multiplicity of ω at time t as the proportion of particles that are exactly at the same place as ω at time t.

Definition 2.3.2. Let $\chi : \Omega \times \mathbb{R}^+ \to X$ be a traffic plan. We define the *path class of* $(\omega, t) \in \Omega \times \mathbb{R}$ *in* χ as the set

$$[\omega, t]_\chi := \{\omega' : \chi(\omega', t) = \chi(\omega, t)\}$$

and the *multiplicity of* χ *at* (ω, t) by $\tilde\theta_\chi(\omega, t) := \mathscr{L}^1([\omega, t]_\chi)$.

Definition 2.3.3. Let $\alpha \in [0, 1]$. We call *dynamical cost* of a traffic plan $\chi : \Omega \times \mathbb{R}^+ \to X$ the functional

$$C^\alpha(\chi) = \int_\Omega \int_{\mathbb{R}^+} \tilde\theta_\chi(\omega, t)^{\alpha-1} |\dot\chi(\omega, t)| dt d\omega, \tag{2.3.2}$$

i.e. $C^\alpha(\chi) = \mathcal{C}(\chi)$ with $c(\chi, t, \omega) = \tilde\theta_\chi(\omega, t)^{\alpha-1}$.

Theorem 2.3.4. *If $\chi_n : \Omega \times \mathbb{R}^+ \to X$ is a sequence in* $\mathrm{TP}(\mu^+, \mu^-)$ *converging to the traffic plan χ, then*

$$C^\alpha(\chi) \le \liminf_n C^\alpha(\chi_n).$$

Proof. Let us denote

$$\delta(x,y) = \begin{cases} 0 & \text{if } x \neq y, \\ 1 & \text{if } x = y. \end{cases}$$

Setting

$$c(\chi,\omega,t) := \left[\int_\Omega \delta(\chi(\omega,t),\chi(\omega',t))\,d\omega'\right]^{\alpha-1},$$

where $\alpha \in [0,1]$, we observe that $c(\chi,t,\omega) = \tilde{\theta}_\chi(\omega,t)^{\alpha-1}$, so that $C^\alpha(\chi) = \mathcal{C}(\chi)$ as defined by (2.3.1). Let us consider a sequence of traffic plans χ_n fiber converging to χ, and $t_n \to t$. We remark that the function

$$\mathbb{R}^N \times \mathbb{R}^N \ni (x,y) \mapsto \delta(x,y) \in \mathbb{R}$$

is upper semicontinuous. Therefore, since $\chi_n(\omega)$ is a 1-Lipschitz curve, if $\chi_n(\omega) \to \chi(\omega)$ and $t_n \to t$ we have

$$\limsup_n \delta(\chi_n(\omega,t_n),\chi_n(\omega',t_n)) \le \delta(\chi(\omega,t),\chi(\omega',t)).$$

Thus, by Fatou lemma,

$$\limsup_n \int_\Omega \delta(\chi_n(\omega,t_n),\chi_n(\omega',t_n))\,d\omega' \le \int_\Omega \delta(\chi(\omega,t),\chi(\omega',t))\,d\omega',$$

and since $\alpha \le 1$,

$$\liminf_n c(\chi_n,\omega,t_n) \ge c(\chi,\omega,t). \tag{2.3.3}$$

Therefore Proposition 2.3.1 ensures that C^α is lower semicontinuous. □

Remark 2.3.5. *It is not difficult to prove the upper semicontinuity of the multiplicity $\theta_\chi(\chi(\omega,t))$, so that the elementary cost $c(\chi,\omega,t) = \theta_\chi(\chi(\omega,t))^{\alpha-1}$ satisfies the hypothesis of Proposition 2.3.1. This yields a new simple proof of Theorem 2.2.9.*

Like in the last paragraph of Section 2.2, it is possible to consider a dynamical cost $C^\alpha_{\mathbb{R}}(\chi)$ for $\chi \in \mathrm{TP}_{\mathbb{R}}(\mu^+,\mu^-)$. Proposition 2.3.1 and Theorem 2.3.4 hold with $\mathrm{TP}_{\mathbb{R}}$ and $C^\alpha_{\mathbb{R}}$ in place of TP and C^α. The compactness of TP_C stated in Proposition 2.2.8 yields:

Proposition 2.3.6. *Let μ^+ and μ^- be probability measures on X, and let $C > 0$ be such that $\mathrm{TP}_C(\mu^+,\mu^-)$ is not empty (for example, take $C \ge \mathrm{diam}(X)$). Then, there exist C^α-minimizers (resp. $C^\alpha_{\mathbb{R}}$-minimizers) in TP_C (resp. $\mathrm{TP}_{\mathbb{R},C}$).*

The argument used to prove Corollary 2.2.10 (that states the existence of E^α-minimizers in TP) is not adaptable to the case of C^α, since neither $C^\alpha(\chi)$ nor $C^\alpha_{\mathbb{R}}(\chi)$ are invariant by time-reparameterization of χ. In particular, the situation where C^α-minimizers in TP_C change as C increases to $+\infty$ is not excluded (this is not the case for E^α, since by the reparameterization argument used to prove Corollary 2.2.10 we know that all minimizers are in TP_C for $C = E^\alpha(\mu^+, \mu^-)$). However, we shall see in Section 2.5, that by using synchronization techniques developed in Section 2.4 we are still able to prove existence of C^α-minimizers in $\mathrm{TP}(\mu^+, \mu^-)$ provided that μ^+ is finite atomic, and of $C^\alpha_{\mathbb{R}}$-minimizers in $\mathrm{TP}_{\mathbb{R}}(\mu^+, \mu^-)$.

2.4 Synchronizable traffic plans

Let us define the support of a traffic plan χ as the set of points with positive multiplicity. This set will be denoted by S_χ.

Definition 2.4.1. *A traffic plan $\chi \in \mathrm{TP}_{\mathbb{R}}(\mu^+, \mu^-)$ (resp. $\mathrm{TP}(\mu^+, \mu^-)$) is said to be* synchronized *(resp.* positive synchronized*) if it is loop-free, and for all x in the support of χ there is a time $t_\chi(x)$ such that $\chi(\omega, t_\chi(x)) = x$ for all $\omega \in \Omega_x$ (i.e. all fibers which pass through x have to pass at the same time).*

Given two traffic plans χ and $\tilde{\chi}$, we say that $\tilde{\chi}$ is a reparameterization of χ if, for almost every $\omega \in \Omega$, the curve $\tilde{\chi}(\omega, \cdot)$ is a reparameterization of $\chi(\omega, \cdot)$. We will say that $\tilde{\chi}$ is an arc length parameterization of χ if, for almost every ω, $\tilde{\chi}(\omega, \cdot)$ is an arc length parameterization of $\chi(\omega, t)$.

Definition 2.4.2. *A traffic plan $\chi \in \mathrm{TP}_{\mathbb{R}}(\mu^+, \mu^-)$ is said to be* synchronizable *(resp.* positive synchronizable*) if there is some reparameterization $\tilde{\chi} \in \mathrm{TP}_{\mathbb{R}}(\mu^+, \mu^-)$ (resp. in $\mathrm{TP}(\mu^+, \mu^-)$) of χ such that $\tilde{\chi}$ is synchronized (resp. positive synchronized).*

Since $\theta_\chi^{\alpha-1} \leq \tilde{\theta}_\chi^{\alpha-1}$ with equality if χ is (positive) synchronized, one can easily deduce that if a traffic plan is synchronized (resp. positive synchronized), then $E^\alpha_{\mathbb{R}}(\chi) = C^\alpha_{\mathbb{R}}(\chi)$ (resp. $E^\alpha(\chi) = C^\alpha(\chi)$).

The aim of this section is to prove that E^α-optimal traffic plans are synchronizable. Indeed, optimal traffic plans are such that there is a finite or countable set of points (x_i) and sets $\Omega_i \subset \Omega_{x_i}$ that form an (almost-)partition of Ω. This fact makes it possible to synchronize independently each tree going through some x_i, and then harmonize globally these synchronizations thanks to the so-called strict single oriented path property that we now discuss.

The strict single path definition was introduced in [26]. Following these authors, a traffic plan is said to be strict single path if all fibers going through x and y have to coincide between x and y. In other terms there is a *single path* (or none) between any two points of the irrigation network. All optimal traffic plans can then be proven to be strict single path up to the removal of a set of fibers with null measure. For our synchronization purposes, we need to use a slight refinement of this notion, namely what we call the strict single oriented path property. To state this property in precise terms, we first need to introduce some definitions.

Definition 2.4.3. *Let χ be a loop-free traffic plan, and define $t_x(\omega) := \inf\{t : \chi(\omega, t) = x\}$. Let x, y in S_χ, and define*

$$\Omega_{\overrightarrow{xy}} := \{\omega \in \Omega_x^\chi \cap \Omega_y^\chi : t_x(\omega) < t_y(\omega)\},$$

the set of fibers passing through x and then through y. We denote by χ_{xy} the restriction of χ to $\cup_{\omega \in \Omega_{\overrightarrow{xy}}} \{\omega\} \times [t_x(\omega), t_y(\omega)]$. It is the traffic plan made of all pieces of fibers of χ joining x to y. Denote its support by $\Gamma^{xy} := S_{\chi_{xy}}$.

Definition 2.4.4. *A traffic plan χ has the* strict single oriented path property *(and we say that χ is* strict single oriented*) if, for every pair x, y such that $|\Omega_{\overrightarrow{xy}}| > 0$, all fibers in $\Omega_{\overrightarrow{xy}}$ coincide between x and y with an arc Γ^{xy} joining x to y, and $\Omega_{\overrightarrow{yx}} = \emptyset$.*

By an immediate adaptation of the strict single path property of optimal traffic plans proven in [26], we have the following result.

Proposition 2.4.5. (Strict single oriented path property) *Let $\alpha \in [0, 1)$ and χ be an optimal traffic plan such that $E^\alpha(\chi) < \infty$. Then, up to removing a zero measure set of fibers, χ has the strict single oriented path property.*

We can now detail the lemmas useful to the prove the synchronizability of E^α-optima.

Lemma 2.4.6. *If χ is strict single oriented and $\tilde{\Omega}_x \subset \Omega_x$, then $\chi_x := \chi \llcorner \tilde{\Omega}_x$ is synchronizable.*

Proof. Let $\tilde{\chi}_x(\omega, t)$ be an arc length parameterization of $\chi_x(\omega, t)$ such that $\tilde{\chi}_x(\omega, 0) = x$. Since $\chi_x(\omega, \cdot)$ is injective, there is only one such parameterization. Let us now prove that $\tilde{\chi}_x$ is synchronized. Indeed, let us consider a point y in the image of χ. Since χ is strict single oriented, there is only one path that connects x to y on the support of the traffic plan χ_x. This allows to define $l_{\chi_x}(y)$ as the distance from x to y (through the support of χ). Since $\tilde{\chi}_x(\omega, \cdot)$ is parameterized by its arc length, we notice that for all $\omega \in \Omega_y \cap \tilde{\Omega}_x$ $\tilde{\chi}_x(\omega, l_\chi(y)) = y$, i.e. $\tilde{\chi}_x$ is synchronized. □

Lemma 2.4.7. *Let χ_1 and χ_2 be synchronized, connected, arc length parameterized, and such that $\chi_1 \cup \chi_2$ is strict single oriented. Then $\chi_1 \cup \chi_2$ is synchronizable.*

Proof. If the supports of χ_1 and χ_2 are disjoints, then $\chi_1 \cup \chi_2$ is already synchronized. Otherwise, let x be a point in the support of both χ_1 and χ_2. Since χ_1 is synchronized, there is some $t_{\chi_1}(x)$ such that for all $\omega \in \Omega_x^{\chi_1}$, $\chi_1(\omega, t_{\chi_1}(x)) = x$. We define $t_{\chi_2}(x)$ analogously. Let us prove that $t_{\chi_1}(x) - t_{\chi_2}(x)$ does not depend on the point x. Let us consider x_1 and x_2 points in the supports of both χ_1 and χ_2. By connectedness and the strict single oriented path property, there is a unique path on the support of χ_1 connecting x_1 and x_2 (the same holds for χ_2). Since χ_1 is arc length parameterized, $t_{\chi_1}(x_1) - t_{\chi_1}(x_2)$ is exactly the distance between x_1 and x_2 (or its opposite, depending on the orientation of the path). Since $\chi_1 \cup \chi_2$ is strict single oriented, the unique path defined by χ_2 is the same as the one of χ_1 so that we have:

$$t_{\chi_1}(x_1) - t_{\chi_1}(x_2) = t_{\chi_2}(x_1) - t_{\chi_2}(x_2).$$

Thus, shifting the time parameterization of χ_2 by $t_{\chi_1}(x) - t_{\chi_2}(x)$ defines a traffic plan $\tilde{\chi}_2$ such that $\chi_1 \cup \tilde{\chi}_2$ is synchronized. □

Definition 2.4.8. *We shall say that a traffic plan χ is finitely (resp. countably) decomposable if there is a finite (resp. countable) set of points (x_i) and sets $\Omega_i \subset \Omega_{x_i}$ that form a partition of Ω (almost everywhere).*

Proposition 2.4.9. *If χ is a strict single oriented countably decomposable traffic plan, then it is synchronizable.*

Proof. Let $\Omega_i \subset \Omega_{x_i}$ defining a countable decomposition of χ, and let us denote $\chi_i := \chi \llcorner \Omega_i$. Lemma 2.4.6 ensures that all the χ_i are synchronizable, and we denote by $\tilde{\chi}_i$ an equivalent synchronized traffic plans. Since χ is strict single oriented, $\cup_i \tilde{\chi}_i$ is strict single oriented. Thus, by induction, the repeated application of Lemma 2.4.7 allows to define a synchronized traffic plan $\tilde{\chi}$ that is the union of time shifted versions of $\tilde{\chi}_i$. Such a traffic plan $\tilde{\chi}$ is a time-reparameterization of χ that is synchronized. □

Proposition 2.4.10. *If μ^+ is finite atomic, then any optimal traffic plan $\chi \in \mathrm{TP}_{\mathbb{R}}(\mu^+, \mu^-)$ is positive synchronizable.*

Proof. Let $(x_i)_{i=1}^n$ be a finite sequence such that $\mu^+ := \sum_{i=1}^n a_i \delta_{x_i}$. The sets defined by $\Omega_1 := \Omega_{x_1}$ and $\Omega_i := \Omega_{x_i} \setminus (\cup_{j<i} \Omega_j)$ for $i > 1$, yield a partition of Ω, so that χ is finitely decomposable. Since χ is optimal, it is strict single oriented, and Proposition 2.4.9 ensures that χ is synchronizable. Since μ^+ is atomic, by the construction of the reparameterization given in Lemma 2.4.7 it is simple to see that, with a suitable time shifting, χ is also positive synchronizable. □

Proposition 2.4.11. *Any optimal traffic plan $\chi \in \mathrm{TP}_{\mathbb{R}}$ is synchronizable.*

Proof. Any optimal traffic plan is countably decomposable (see [26, Lemma 3.11]) and strict single oriented. Thus, by Proposition 2.4.9, it is synchronizable. □

2.5 Equivalence of the dynamical and classical irrigation problems

In the same way as for E^α, we define:

$$C^\alpha(\mu^+,\mu^-) := \inf_{\mathrm{TP}(\mu^+,\mu^-)} C^\alpha(\chi), \qquad C^\alpha_{\mathbb{R}}(\mu^+,\mu^-) := \inf_{\mathrm{TP}_{\mathbb{R}}(\mu^+,\mu^-)} C^\alpha_{\mathbb{R}}(\chi).$$

Theorem 2.5.1. *If μ^+ is finite atomic, then, for all $\alpha \in [0,1]$,*

$$E^\alpha(\mu^+,\mu^-) = C^\alpha(\mu^+,\mu^-),$$

and

$$E^\alpha_{\mathbb{R}}(\mu^+,\mu^-) = C^\alpha_{\mathbb{R}}(\mu^+,\mu^-).$$

Proof. We remark that, by the definition of E^α and C^α, we immediately have the inequality

$$E^\alpha(\chi) \le C^\alpha(\chi) \quad \text{for all traffic plan } \chi, \tag{2.5.1}$$

so that,

$$E^\alpha(\mu^+,\mu^-) \le C^\alpha(\mu^+,\mu^-) \quad \forall \alpha \in [0,1].$$

Let χ be a minimizer of E^α. Proposition 2.4.10 ensures that there is a reparameterization $\tilde{\chi}$ of χ such that $\tilde{\chi}$ is positive synchronized, so that

$$E^\alpha(\mu^+,\mu^-) = E^\alpha(\chi) = E^\alpha(\tilde{\chi}) = C^\alpha(\tilde{\chi}).$$

Thus, $E^\alpha(\mu^+,\mu^-) = C^\alpha(\mu^+,\mu^-)$ for all $\alpha \in [0,1]$. Finally, Proposition 2.4.11 yields $E^\alpha_{\mathbb{R}}(\mu^+,\mu^-) = C^\alpha_{\mathbb{R}}(\mu^+,\mu^-)$ for all $\alpha \in [0,1]$. □

By Proposition 2.4.11, we also have:

Theorem 2.5.2. *Let μ^+ and μ^- be two probability measures. Then*

$$E^\alpha_{\mathbb{R}}(\mu^+,\mu^-) = C^\alpha_{\mathbb{R}}(\mu^+,\mu^-).$$

Theorem 2.5.2 states the equivalence of the cost given by the dynamical and the classical irrigation problem. Concerning minimizers, we can observe as a direct consequence of Theorem 2.5.2 and (2.5.1) that every $C^\alpha_{\mathbb{R}}$-minimizer is an $E^\alpha_{\mathbb{R}}$-minimizer. Conversely, by Proposition 2.4.11, any $E^\alpha_{\mathbb{R}}$-minimizer can be reparameterized so that it gives a $C^\alpha_{\mathbb{R}}$-minimizer. The same considerations are true for E^α and C^α if μ^+ is finite atomic thanks to Proposition 2.4.10. Thus, in both these cases, the extended dynamical and classical irrigation problems yield exactly the same minimizers (up to reparameterization).

In particular, as a by-product, we obtain the existence of C^α-minimizers if μ^+ is finite atomic, and existence of $C^\alpha_{\mathbb{R}}$-minimizers in general.

As a particular consequence of the fact that every $C^\alpha_{\mathbb{R}}$-minimizer is an $E^\alpha_{\mathbb{R}}$-minimizer, we notice that $C^\alpha_{\mathbb{R}}$-minimizers inherit all the regularity properties of $E^\alpha_{\mathbb{R}}$-minimizers (the same holds for C^α-minimizers, in the case μ^+ is finite atomic). Thus we can translate the regularity results in Section 2.2 in the $C^\alpha_{\mathbb{R}}$ framework.

Proposition 2.5.3. *Let $\alpha \in [0,1]$, μ^+ and μ^- be finite atomic measures, and $\chi \in \mathrm{TP}(\mu^+,\mu^-)$ be a C^α-minimizer. Then χ is a finite tree made of segments.*

Theorem 2.5.4. *Let $\alpha \in (1-\frac{1}{N},1)$, μ^+ and μ^- be probability measures, and $\chi \in \mathrm{TP}_{\mathbb{R}}(\mu^+,\mu^-)$ be a $C^\alpha_{\mathbb{R}}$-minimizer. Assume that the supports of μ^+ and μ^- are at positive distance. In any closed ball $B(x,r)$ not meeting the supports of μ^+ and μ^-, the traffic plan χ has the structure of a finite graph.*

2.6 Stability with respect to the cost

In this section we study the regularity with respect to α of $E^\alpha(\mu^+,\mu^-)$ for fixed μ^+ and μ^-. By the equivalence of $E^\alpha_{\mathbb{R}}$ and $C^\alpha_{\mathbb{R}}$ (Theorem 2.5.2), and E^α and C^α when μ^+ is finite atomic (Theorem 2.5.1), one can deduce similar stability results for the dynamical cost.

We start studying the regularity with respect to α of $E^\alpha(\chi)$ for a fixed traffic plan χ.

Lemma 2.6.1. *Let χ be a traffic plan. Then $[0,1] \ni \alpha \mapsto E^\alpha(\chi) \in \mathbb{R}^+ \cup \{+\infty\}$ is non-increasing. Fix now $\alpha \in [0,1)$. Then:*

(i) If $E^\alpha(\chi) < +\infty$, then $\beta \mapsto E^\beta(\chi)$ is finite and continuous on $[\alpha,1]$.

(ii) If $E^\alpha(\chi) = +\infty$, then $E^{\alpha_n}(\chi) \to +\infty$ for any decreasing sequence $\alpha_n \searrow \alpha$.

Proof. The monotonicity of $\alpha \mapsto E^\alpha(\chi)$ is trivial.

Let χ be such that $E^\alpha(\chi) < +\infty$ and let $\beta_n \in [\alpha,1]$ such that $\beta_n \to \beta$. For all $(\omega,t) \in \Omega \times \mathbb{R}^+$, we have

$$\theta_\chi(\chi(\omega,t))^{\beta_n-1} \to \theta_\chi(\chi(\omega,t))^{\beta-1}.$$

In addition, as $\theta_\chi(\chi(\omega,t)) \le 1$, we have

$$0 \le \theta_\chi(\chi(\omega,t))^{\beta_n-1} \le \theta_\chi(\chi(\omega,t))^{\alpha-1}$$

Thus, since

$$E^\alpha(\chi) = \int_\Omega \int_{\mathbb{R}^+} \theta_\chi(\chi(\omega,t))^{\alpha-1} |\dot\chi(\omega,t)|\, dt\, d\omega < \infty,$$

the dominated convergence theorem ensures the convergence of $E^{\beta_n}(\chi)$ to $E^{\beta}(\chi)$.

Let us now consider a traffic plan χ such that $E^{\alpha}(\chi) = +\infty$, and let α_n be a decreasing sequence converging to α. Then for all $(\omega, t) \in \Omega \times \mathbb{R}^+$, $\theta_\chi(\chi(\omega,t))^{\alpha_n - 1}$ is increasingly converging to $\theta_\chi(\chi(\omega,t))^{\alpha-1}$. Thus, the monotone convergence theorem ensures that $E^{\alpha_n}(\chi) \to +\infty$. □

Now we can study the stability of $E^{\alpha}(\mu^+, \mu^-)$ with respect to α.

Proposition 2.6.2. *Let μ^+ and μ^- be two probability measures. The function $[0,1] \ni \alpha \mapsto E^{\alpha}(\mu^+, \mu^-) \in \mathbb{R}^+ \cup \{+\infty\}$ is non-increasing, right continuous and left lower semi-continuous.*

Proof. For simplicity of notation set $f(\alpha) := E^{\alpha}(\mu^+, \mu^-)$. Observe that, since $\alpha \mapsto E^{\alpha}(\chi)$ is non-increasing for all χ, f is non-increasing being an infimum of non-increasing functions. Thus, f is left lower semicontinuous, i.e.

$$\liminf_n f(\alpha_n) \geq f(\alpha) \quad \text{for all } \alpha_n \nearrow \alpha.$$

In what follows, χ_β will always denote an optimal traffic plan for the exponent β, i.e. such that $E^{\beta}(\chi_\beta) = f(\beta)$. Let us consider a decreasing sequence α_n such that $\alpha_n \searrow \alpha$ and a sequence of optimal traffic plans χ_{α_n}.

By Lemma 2.6.1 and the optimality of χ_{α_n} for E^{α_n} we get

$$f(\alpha) = E^{\alpha}(\chi_\alpha) = \lim_n E^{\alpha_n}(\chi_\alpha) \geq \limsup_n E^{\alpha_n}(\chi_{\alpha_n}) \geq \liminf_n E^{\alpha_n}(\chi_{\alpha_n}). \tag{2.6.1}$$

If $\liminf_n E^{\alpha_n}(\chi_{\alpha_n}) = +\infty$, there is nothing to prove. Otherwise, up to apply the reparameterization argument used to prove Corollary 2.2.9, we can assume that $\chi_{\alpha_n} \in \mathrm{TP}_C$ for some $C > 0$. Thus, by Proposition 2.2.8, there is a subsequence $\chi_{\alpha_{n_k}}$ such that

$$\chi_{\alpha_{n_k}} \to \chi \quad \text{and} \quad \liminf_k E^{\alpha_{n_k}}(\chi_{\alpha_{n_k}}) = \liminf_n E^{\alpha_n}(\chi_{\alpha_n}). \tag{2.6.2}$$

Recalling that $\alpha \mapsto E^{\alpha}(\chi)$ is non-increasing, and that E^{α_m} is lower semicontinuous for m fixed, we have

$$\liminf_k E^{\alpha_{n_k}}(\chi_{\alpha_{n_k}}) \geq \liminf_k E^{\alpha_m}(\chi_{\alpha_{n_k}}) \geq E^{\alpha_m}(\chi) \quad \text{for all } m. \tag{2.6.3}$$

By Lemma 2.6.1, $\lim_m E^{\alpha_m}(\chi) = E^{\alpha}(\chi)$ so that (2.6.1), (2.6.2) and (2.6.3) yield

$$f(\alpha) \geq \limsup_n f(\alpha_n) \geq \liminf_n f(\alpha_n) \geq E^{\alpha}(\chi) \geq f(\alpha).$$

□

Corollary 2.6.3. *Let $\alpha_n \in [0,1]$ be a decreasing sequence converging to α, and let μ^+ and μ^- be two probability measures. If χ_{α_n} are optimal traffic plans for E^{α_n} and $\chi_{\alpha_n} \to \chi$, then χ is optimal for E^α.*

Proof. By Proposition 2.6.2, and since $\alpha \mapsto E^\alpha(\chi)$ is non-increasing and E^{α_m} is lower semicontinuous for fixed m, we have

$$E^\alpha(\mu^+,\mu^-) = \lim_n E^{\alpha_n}(\chi_{\alpha_n}) \geq \liminf_n E^{\alpha_m}(\chi_{\alpha_n}) \geq E^{\alpha_m}(\chi).$$

Since by Lemma 2.6.1 $\lim_m E^{\alpha_m}(\chi) = E^\alpha(\chi)$, χ is optimal. □

If we now constrain α to be in $(1-\frac{1}{N},1]$, we are able to say more. Indeed, in this case, Proposition 2.2.14 allows us to approximate μ^+ and μ^- with atomic measures μ_n^+ and μ_n^- in such a way that $E^\alpha(\mu^+,\mu^-)$ is a uniform limit (locally in α) of $E^\alpha(\mu_n^+,\mu_n^-)$. Then it is sufficient to prove that $E^\alpha(\mu_n^+,\mu_n^-)$ is continuous for any n, in order to have that $E^\alpha(\mu^+,\mu^-)$ is continuous on $(1-\frac{1}{N},1]$.

Lemma 2.6.4. *Let $\mu^+ = \sum_{i=1}^{k_1} a_i\delta_{x_i}$ and $\mu^- = \sum_{i=1}^{k_2} b_i\delta_{y_i}$ be atomic measures such that $\sum_{i=1}^{k_1} a_i = \sum_{i=1}^{k_2} b_i$ (the irrigating and the irrigated measure have the same mass). Then $\alpha \mapsto E^\alpha(\mu^+,\mu^-)$ is continuous on $[0,1]$.*

Proof. By Proposition 2.2.17, we know that, for all $\alpha \in [0,1]$, an optimum for the irrigation problem can be viewed as a weighted and oriented finite graph G. Then, if we call χ_α an optimum for E^α, we have

$$E^\alpha(\mu^+,\mu^-) = E^\alpha(\chi_\alpha) = \sum_{i=1}^{n_\alpha} l_i m_i^\alpha,$$

where the l_i and m_i are respectively the lengths and weigths of the edges of G. Then, since

$$\beta \mapsto E^\beta(\chi_\alpha) = \sum_{i=1}^{n_\alpha} l_i m_i^\beta$$

is continuous and finite on $[0,1]$, we see that $E^\alpha(\mu^+,\mu^-)$ is finite on $[0,1]$. Moreover we already know by Proposition 2.6.2 that $E^\alpha(\mu^+,\mu^-)$ is left lower semicontinuous and right continuous. So, in order to conclude it is sufficient to prove that $E^\alpha(\mu^+,\mu^-)$ is left upper semicontinuous. Let (α_n) be a sequence such that $\alpha_n \nearrow \alpha$. The continuity of $\beta \mapsto E^\beta(\chi_\alpha)$ ensures that

$$\limsup_n E^{\alpha_n}(\mu^+,\mu^-) = \limsup_n E^{\alpha_n}(\chi_{\alpha_n}) \leq \limsup_n E^{\alpha_n}(\chi_\alpha) = E^\alpha(\chi_\alpha) = E^\alpha(\mu^+,\mu^-).$$

□

Theorem 2.6.5. *Let $\alpha_n \in [1-\frac{1}{N}, 1]$ be a sequence converging to α. If the traffic plans χ_{α_n} are optimal for E^{α_n} and $\chi_{\alpha_n} \to \chi$, then χ is optimal for E^{α}.*

Proof. By Proposition 2.2.14, for all μ^+ and μ^- there are atomic measures μ_n^+ and μ_n^- such that $E^\alpha(\mu_n^+, \mu_n^-)$ converges uniformly to $E^\alpha(\mu^+, \mu^-)$ on $(1-\frac{1}{N}, 1]$. Lemma 2.6.4 asserts that $\alpha \mapsto E^\alpha(\mu_n^+, \mu_n^-)$ is continuous, so that $\alpha \mapsto E^\alpha(\mu^+, \mu^-)$ is continuous on $(1-\frac{1}{N}, 1]$. By the same kind of argument as in the proof of Corollary 2.6.3, we deduce that χ is optimal. If $\alpha = 1-\frac{1}{N}$, we can suppose that up to a subsequence $\alpha_n \searrow \alpha$, so that Corollary 2.6.3 ensures that χ is optimal (possibly trivially optimal in the case $E^\alpha(\chi) = \infty$). □

Remark 2.6.6. *In the case $\alpha = 1$, the irrigation problem for the cost E^α is equivalent to the classical Monge-Kantorovich problem (see [110, 85, 132]). For that particular case, Theorem 2.6.5 ensures that the transference plan associated to a sequence of optimal traffic plans χ_{α_n}, where $\alpha_n \to 1$, converges, up to a subsequence, to an optimal transference plan for the Monge-Kantorovich problem.*

Chapter 3

Variational models for the incompressible Euler equations

3.1 Introduction

[1] The velocity of an incompressible fluid moving inside a region D is mathematically described by a time-dependent and divergence-free vector field $\boldsymbol{u}(t,x)$ which is parallel to the boundary ∂D. The Euler equations for incompressible fluids describes the evolution of such velocity field $\boldsymbol{u}$ in terms of the pressure field p:

$$\begin{cases} \partial_t \boldsymbol{u} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p & \text{in } [0,T] \times D, \\ \operatorname{div} \boldsymbol{u} = 0 & \text{in } [0,T] \times D, \\ \boldsymbol{u} \cdot n = 0 & \text{on } [0,T] \times \partial D. \end{cases} \tag{3.1.1}$$

Let us assume that $\boldsymbol{u}$ is smooth, so that it produces a unique flow g, given by

$$\begin{cases} \dot{g}(t,a) = \boldsymbol{u}(t, g(t,a)), \\ g(0,a) = a. \end{cases}$$

By the incompressibility condition, we get that at each time t the map $g(t,\cdot) : D \to D$ is a measure-preserving diffeomorphism of D, that is

$$g(t,\cdot)_\# \mu_D = \mu_D,$$

(here and in the sequel $f_\# \mu$ is the push-forward of a measure μ through a map f, and μ_D is the volume measure of the manifold D). Writing Euler equations in terms of g, we

[1] This chapter is based on two joint works with Luigi Ambrosio [8, 9].

get

$$\begin{cases} \ddot{g}(t,a) = -\nabla p\,(t, g(t,a)) & (t,a) \in [0,T] \times D, \\ g(0,a) = a & a \in D, \\ g(t,\cdot) \in \mathrm{SDiff}(D) & t \in [0,T]. \end{cases} \tag{3.1.2}$$

Viewing the space $\mathrm{SDiff}(D)$ of measure-preserving diffeomorphisms of D as an infinite-dimensional manifold with the metric inherited from the embedding in L^2, and with tangent space made by the divergence-free vector fields, Arnold interpreted the equation above, and therefore (3.1.1), as a *geodesic* equation on $\mathrm{SDiff}(D)$ [15]. According to this intepretation, one can look for solutions of (3.1.2) by minimizing

$$T \int_0^T \int_D \frac{1}{2} |\dot{g}(t,x)|^2 \, d\mu_D(x) \, dt \tag{3.1.3}$$

among all paths $g(t,\cdot) : [0,T] \to \mathrm{SDiff}(D)$ with $g(0,\cdot) = f$ and $g(T,\cdot) = h$ prescribed (typically, by right invariance, f is taken as the identity map $\boldsymbol{i}$), and the pressure field arises as a Lagrange multiplier from the incompressibility constraint (the factor T in front of the integral is just to make the functional scale invariant in time). We shall denote by $\delta(f,h)$ the Arnold distance in $\mathrm{SDiff}(D)$, whose square is defined by the above-mentioned variational problem in the time interval $[0,1]$.

Although in the traditional approach to (3.1.1) the initial velocity is prescribed, while in the minimization of (3.1.3) is not, this variational problem has an independent interest and leads to deep mathematical questions, namely existence of relaxed solutions, gap phenomena and necessary and sufficient optimality conditions, that are investigated in this chapter. We also remark that *no* existence result of distributional solutions of (3.1.1) is known when $d > 2$ (the case $d = 2$ is different, thanks to the vorticity formulation of (3.1.1)), see [94], [36] for a discussion on this topic and other concepts of weak solutions to (3.1.1).

On the positive side, Ebin and Marsden proved in [58] that, when D is a smooth compact manifold with no boundary, the minimization of (3.1.3) leads to a unique solution, corresponding also to a solution to Euler equations, if f and h are sufficiently close in a suitable Sobolev norm.

On the negative side, Shnirelman proved in [121], [122] that when $d \geq 3$ the infimum is not attained in general, and that when $d = 2$ there exists $h \in \mathrm{SDiff}(D)$ which cannot be connected to $\boldsymbol{i}$ by a path with finite action. These "negative" results motivate the study of relaxed versions of Arnold's problem.

The first relaxed version of Arnold's minimization problem was introduced by Brenier in [31]: he considered probability measures $\boldsymbol{\eta}$ in $\Omega(D)$, the space of continuous paths $\omega : [0,T] \to D$, and minimized the energy

$$\mathscr{A}_T(\boldsymbol{\eta}) := T \int_{\Omega(D)} \int_0^T \frac{1}{2} |\dot{\omega}(\tau)|^2 \, d\tau \, d\boldsymbol{\eta}(\omega),$$

with the constraints

$$(e_0, e_T)_\# \boldsymbol{\eta} = (\boldsymbol{i}, h)_\# \mu_D, \qquad (e_t)_\# \boldsymbol{\eta} = \mu_D \quad \forall t \in [0, T] \tag{3.1.4}$$

(here and in the sequel $e_t(\omega) := \omega(t)$ are the evaluation maps at time t). According to Brenier, we shall call these $\boldsymbol{\eta}$ *generalized incompressible flows* in $[0, T]$ between $\boldsymbol{i}$ and h. Obviously any sufficiently regular path $g(t, \cdot) : [0, 1] \to S(D)$ induces a generalized incompressible flow $\boldsymbol{\eta} = (\Phi_g)_\# \mu_D$, where $\Phi_g : D \to \Omega(D)$ is given by $\Phi_g(x) = g(\cdot, x)$, but the converse is far from being true: the main difference between classical and generalized flows consists in the fact that fluid paths starting from different points are allowed to cross at a later time, and fluid paths starting from the same point are allowed to split at a later time. This approach is by now quite common, see for instance [4] (DiPerna-Lions theory), [25] (branched optimal transportation), [97], [133].

Brenier's formulation makes sense not only if $h \in \mathrm{SDiff}(D)$, but also when $h \in S(D)$, where $S(D)$ is the space of measure-preserving maps $h : D \to D$, not necessarily invertible or smooth. In the case $D = [0, 1]^d$, existence of admissible paths with finite action connecting $\boldsymbol{i}$ to any $h \in S(D)$ was proved in [31], together with the existence of paths with minimal action. Furthermore, a consistency result was proved: smooth solutions to (3.1.1) are optimal even in the larger class of the generalized incompressible flows, provided the pressure field p satisfies

$$T^2 \sup_{t \in [0,T]} \sup_{x \in D} |\nabla_x^2 p(t, x)| \leq \pi^2, \tag{3.1.5}$$

and are the unique ones if the inequality is strict. When $\boldsymbol{\eta} = (\Phi_g)_\# \mu_D$ we can recover $g(t, \cdot)$ from $\boldsymbol{\eta}$ using the identity

$$(e_0, e_t)_\# \boldsymbol{\eta} = (\boldsymbol{i}, g(t, \cdot))_\# \mu_D, \qquad t \in [0, T].$$

Brenier found in [31] examples of action-minimizing paths $\boldsymbol{\eta}$ (for instance in the unit ball of $\mathbb{R}^2$, between $\boldsymbol{i}$ and $-\boldsymbol{i}$) where no such representation is possible. The same examples show that the upper bound (3.1.5) is sharp. Notice however that $(e_0, e_t)_\# \boldsymbol{\eta}$ is a measure-preserving plan, i.e. a probability measure in $D \times D$ having both marginals equal to μ_D. Denoting by $\Gamma(D)$ the space of measure-preserving plans, it is therefore natural to consider $t \mapsto (e_0, e_t)_\# \boldsymbol{\eta}$ as a "minimizing geodesic" between $\boldsymbol{i}$ and h in the larger space of measure-preserving plans. Then, to be consistent, one has to extend Brenier's minimization problem considering paths connecting $\gamma, \eta \in \Gamma(D)$. We define this extension, that reveals to be useful also to connect this model to the Eulerian-Lagrangian one in [35], and to obtain necessary and sufficient optimality conditions even when only "deterministic" data $\boldsymbol{i}$ and h are considered (because, as we said, the path might be non-deterministic in between). In this presentation of our results, however, to simplify the matter as much as possible, we shall consider the case of paths $\boldsymbol{\eta}$ between $\boldsymbol{i}$ and $h \in S(D)$ only.

In Section 3.5 we study the relation between the relaxation δ_* of the Arnold distance, defined by

$$\delta_*(h) := \inf\left\{\liminf_{n\to\infty} \delta(\boldsymbol{i}, h_n) : \ h_n \in \mathrm{SDiff}(D), \ \int_D |h_n - h|^2 \, d\mu_D \to 0\right\},$$

and the distance $\overline{\delta}(\boldsymbol{i}, h)$ arising from the minimization of the Lagrangian model. It is not hard to show that $\overline{\delta}(\boldsymbol{i}, h) \leq \delta_*(h)$, and a natural question is whether equality holds, or a gap phenomenon occurs. In the case $D = [0,1]^d$ with $d > 2$, an important step forward was obtained by Shnirelman in [122], who proved that equality holds when $h \in \mathrm{SDiff}(D)$; Shnirelman's construction provides an approximation (with convergence of the action) of generalized flows connecting $\boldsymbol{i}$ to h by smooth flows still connecting $\boldsymbol{i}$ to h. The main result of this section is the proof that no gap phenomenon occurs, still in the case $D = [0,1]^d$ with $d > 2$, even when non-deterministic final data (i.e. measure-preserving plans) are considered. The proof of this fact is based on an auxiliary approximation result, Theorem 3.5.3, valid in any number of dimensions, which we believe of independent interest: it allows to approximate, with convergence of the action, any generalized flow $\boldsymbol{\eta}$ in $[0,1]^d$ by $W^{1,2}$ flows (in time) induced by measure-preserving maps $g(t,\cdot)$. This fact shows that the "negative" result of Shnirelman on the existence in dimension 2 of non-attainable diffeomorphisms is due to the regularity assumption on the path, and it is false if one allows for paths in the larger space $S(D)$. The proof of Theorem 3.5.3 uses some key ideas from [122] (in particular the combination of law of large numbers and smoothing of discrete families of trajectories), and some ideas coming from the theory of optimal transportation.

Minimizing generalized paths $\boldsymbol{\eta}$ are not unique in general, as shown in [31]; however, Brenier proved in [33] that the gradient of the pressure field p, identified by the distributional relation

$$\nabla p(t,x) = -\partial_t \overline{\boldsymbol{v}}_t(x) - \mathrm{div}\left(\overline{\boldsymbol{v}\otimes\boldsymbol{v}}_t(x)\right), \tag{3.1.6}$$

is indeed unique. Here $\overline{\boldsymbol{v}}_t(x)$ is the "effective velocity", defined by $(e_t)_\#(\dot\omega(t)\boldsymbol{\eta}) = \overline{\boldsymbol{v}}_t\mu_D$, and $\overline{\boldsymbol{v}\otimes\boldsymbol{v}}_t$ is the quadratic effective velocity, defined by $(e_t)_\#(\dot\omega(t)\otimes\dot\omega(t)\boldsymbol{\eta}) = \overline{\boldsymbol{v}\otimes\boldsymbol{v}}_t\mu_D$. The proof of this fact is based on the so-called dual least action principle: if $\boldsymbol{\eta}$ is optimal, we have

$$\mathscr{A}_T(\boldsymbol{\nu}) \geq \mathscr{A}_T(\boldsymbol{\eta}) + \langle p, \rho^{\boldsymbol{\nu}} - 1\rangle \tag{3.1.7}$$

for any measure $\boldsymbol{\nu}$ in $\Omega(D)$ such that $(e_0, e_T)_\#\boldsymbol{\nu} = (\boldsymbol{i}, h)_\#\mu_D$ and $\|\rho^{\boldsymbol{\nu}} - 1\|_{C^1} \leq 1/2$. Here $\rho^{\boldsymbol{\nu}}$ is the (absolutely continuous) density produced by the flow $\boldsymbol{\nu}$, defined by $\rho^{\boldsymbol{\nu}}(t,\cdot)\mu_D =$

$(e_t)_\#\boldsymbol{\nu}$. In this way, the incompressibility constraint can be slightly relaxed and one can work with the augmented functional (still minimized by $\boldsymbol{\eta}$)

$$\boldsymbol{\nu} \mapsto \mathscr{A}_T(\boldsymbol{\nu}) - \langle p, \rho^{\boldsymbol{\nu}} - 1 \rangle,$$

whose first variation leads to (3.1.6).

In Theorem 3.6.2, still using the key Proposition 2.1 from [33], we provide a simpler proof and a new interpretation of the dual least action principle.

A few years later, Brenier introduced in [35] a new relaxed version of Arnold's problem of a mixed Eulerian-Lagrangian nature: the idea is to add to the Eulerian variable x a Lagrangian one a representing, at least when $f = \boldsymbol{i}$, the initial position of the particle; then, one minimizes a functional of the Eulerian variables (density and velocity), depending also on a. Brenier's motivation for looking at the new model was that this formalism allows to show much stronger regularity results for the pressure field, namely $\partial_{x_i} p$ are locally finite measures in $(0,T) \times D$. In Section 3.3.3 we describe in detail this new model and, in Section 3.4, we show that the two models are basically equivalent. This result will be used by us to transfer the regularity informations on the pressure field up to the Lagrangian model, thus obtaining the validity of (3.1.7) for a much larger class of generalized flows $\boldsymbol{\nu}$, that we call flows with bounded compression. The proof of the equivalence follows by a general principle (Theorem 3.2.4, borrowed from [11]) that allows to move from an Eulerian to a Lagrangian description, lifting solutions to the continuity equation to measures in the space of continuous maps.

In Section 3.6 we look for necessary and sufficient optimality conditions for the geodesic problem. These conditions require that the pressure field p is a function and not only a distribution: this technical result is achieved in the last section, where, by carefully analyzing and improving Brenier's difference-quotient argument, we show that $\partial_{x_i} p \in L^2_{\mathrm{loc}}\big((0,T);\mathcal{M}_{\mathrm{loc}}(D)\big)$ (this implies, by Sobolev embedding, $p \in L^2_{\mathrm{loc}}\big((0,T);L^{d/(d-1)}_{\mathrm{loc}}(D)\big)$).

In this final section, although we do not see a serious obstruction to the extension of our results to a more general framework, we consider the case of the flat torus $\mathbb{T}^d$ only, and we shall denote by $\mu_{\mathbb{T}}$ the canonical measure on the flat torus. We observe that in this case $p \in L^2_{\mathrm{loc}}\big((0,T);L^{d/(d-1)}(\mathbb{T}^d)\big)$ and so, taking into account that the pressure field in (3.1.7) is uniquely determined up to additive time-dependent constants, we may assume that $\int_{\mathbb{T}^d} p(t,\cdot)d\mu_{\mathbb{T}} = 0$ for almost all $t \in (0,T)$.
The first elementary remark is that any integrable function q in $(0,T)\times\mathbb{T}^d$ with $\int_{\mathbb{T}^d} q(t,\cdot)\,d\mu_{\mathbb{T}} = 0$ for almost all $t \in (0,T)$ provides us with a null-lagrangian for the geodesic problem, as the incompressibility constraint gives

$$\int_{\Omega(\mathbb{T}^d)} \int_0^T q(t,\omega(t))\,dt\,d\boldsymbol{\nu}(\omega) = \int_0^T \int_{\mathbb{T}^d} q(t,x)\,d\mu_{\mathbb{T}}(x)\,dt = 0$$

for any generalized incompressible flow $\boldsymbol{\nu}$. Taking also the constraint $(e_0, e_T)_\#\boldsymbol{\nu} = (\boldsymbol{i}, h)_\#\mu$ into account, we get

$$\mathscr{A}_T(\boldsymbol{\nu}) = T\int_{\Omega(\mathbb{T}^d)}\left(\int_0^T \frac{1}{2}|\dot\omega(t)|^2 - q(t,\omega)\,dt\right)d\boldsymbol{\nu}(\omega) \geq \int_{\mathbb{T}^d} c_q^T(x,h(x))\,d\mu_{\mathbb{T}}(x),$$

where $c_q^T(x,y)$ is the minimal cost associated with the Lagrangian $T\int_0^T \frac{1}{2}|\dot\omega(t)|^2 - q(t,\omega)\,dt$. Since this lower bound depends only on h, we obtain that any $\boldsymbol{\eta}$ satisfying (3.1.4) and concentrated on c_q-minimal paths, for some $q \in L^1$, is optimal, and $\overline{\delta}^2(\boldsymbol{i},h) = \int c_q^T(\boldsymbol{i},h)\,d\mu_{\mathbb{T}}$. This is basically the argument used by Brenier in [31] to show the minimality of smooth solutions to (3.1.1), under assumption (3.1.5): indeed, this condition guarantees that solutions of $\ddot\omega(t) = -\nabla p(t,\omega)$ (i.e. stationary paths for the Lagrangian, with $q = p$) are also minimal.

We are able to show that basically this condition is *necessary and sufficient* for optimality if the pressure field is globally integrable (see Theorem 3.6.12). However, since no global in time regularity result for the pressure field is presently known, we have also been looking for necessary and sufficient optimality conditions that don't require the global integrability of the pressure field. Using the regularity $p \in L^1_{\mathrm{loc}}((0,T);L^r(D))$ for some $r > 1$, guaranteed in the case $D = \mathbb{T}^d$ with $r = d/(d-1)$ by the results contained in the last saction, we show in Theorem 3.6.8 that any optimal $\boldsymbol{\eta}$ is concentrated on *locally minimizing* paths for the Lagrangian

$$\mathcal{L}_p(\omega) := \int \frac{1}{2}|\dot\omega(t)|^2 - p(t,\omega)\,dt \tag{3.1.8}$$

Since we are going to integrate p along curves, this statement is not invariant under modifications of p in negligible sets, and the choice of a *specific* representative $\bar p(t,x) := \liminf_{\varepsilon\downarrow 0} p(t,\cdot) * \phi_\varepsilon(x)$ in the Lebesgue equivalence class is needed. Moreover, the necessity of pointwise uniform estimates on p_ε requires the integrability of $Mp(t,x)$, the maximal function of $p(t,\cdot)$ at x (see (3.6.11)).

In addition, we identify a second necessary (and more hidden) optimality condition. In order to state it, let us consider an interval $[s,t] \subset (0,T)$ and the cost function

$$c_p^{s,t}(x,y) := \inf\left\{\int_s^t \frac{1}{2}|\dot\omega(\tau)|^2 - p(\tau,\omega)\,d\tau :\ \omega(s) = x,\, \omega(t) = y,\, Mp(\tau,\omega) \in L^1(s,t)\right\}. \tag{3.1.9}$$

(the assumption $Mp(\tau,\omega) \in L^1(s,t)$ is forced by technical reasons). Recall that, according to the theory of optimal transportation, a probability measure λ in $\mathbb{T}^d \times \mathbb{T}^d$ is said to be c-optimal if

$$\int_{\mathbb{T}^d\times\mathbb{T}^d} c(x,y)\,d\lambda' \geq \int_{\mathbb{T}^d\times\mathbb{T}^d} c(x,y)\,d\lambda$$

for any probability measure λ' having the same marginals μ_1, μ_2 of λ. We shall also denote $W_c(\mu_1, \mu_2)$ the minimal value, i.e. $\int_{\mathbb{T}^d\times\mathbb{T}^d} c\, d\lambda$, with λ c-optimal. Now, let $\boldsymbol{\eta}$ be an optimal generalized incompressible flow between $\boldsymbol{i}$ and h; according to the disintegration theorem, we can represent $\boldsymbol{\eta} = \int \boldsymbol{\eta}_a\, d\mu_D(a)$, with $\boldsymbol{\eta}_a$ concentrated on curves starting at a (and ending, since our final conditions is deterministic, at $h(a)$), and consider the plans $\lambda_a^{s,t} = (e_s, e_t)_\# \boldsymbol{\eta}_a$. We show that

$$\text{for all } [s,t] \subset (0,T), \qquad \lambda_a^{s,t} \text{ is } c_p^{s,t}\text{-optimal for } \mu_{\mathbb{T}}\text{-a.e. } a \in \mathbb{T}^d. \tag{3.1.10}$$

Roughly speaking, this condition tells us that one has not only to move mass from x to y achieving $c_p^{s,t}$, but also to optimize the distribution of mass between time s and time t. In the "deterministic" case when either $(e_0, e_s)_\# \boldsymbol{\eta}$ or $(e_0, e_t)_\# \boldsymbol{\eta}$ are induced by a transport map g, the plan $\lambda_a^{s,t}$ has $\delta_{g(a)}$ either as first or as second marginal, and therefore it is uniquely determined by its marginals (it is indeed the product of them). This is the reason why condition (3.1.10) does not show up in the deterministic case.

Finally, we show in Theorem 3.6.12 that the two conditions are also sufficient, even on general manifolds D: if, for some $r > 1$ and $q \in L^1_{\mathrm{loc}}\left((0,T); L^r(D)\right)$, a generalized incompressible flow $\boldsymbol{\eta}$ concentrated on locally minimizing curves for the Lagrangian $\mathcal{L}_q$ satisfies

$$\text{for all } [s,t] \subset (0,T), \qquad \lambda_a^{s,t} \text{ is } c_q^{s,t}\text{-optimal for } \mu_D\text{-a.e. } a \in D,$$

then $\boldsymbol{\eta}$ is optimal in $[0,T]$, and q is the pressure field.

These results show a somehow unexpected connection between the variational theory of incompressible flows and the theory developed by Bernard-Buffoni [20] of measures in the space of action-minimizing curves; in this framework one can fit Mather's theory as well as optimal transportation problems on manifolds, with a geometric cost. In our case the only difference is that the Lagrangian is possibly nonsmoooth (but hopefully not so bad), and not given *a priori*, but generated by the problem itself. Our approach also yields (see Corollary 3.6.13) a new variational characterization of the pressure field, as a maximizer of the family of functionals (for $[s,t] \subset (0,T)$)

$$q \mapsto \int_{\mathbb{T}^d} W_{c_q^{s,t}}(\eta_a^s, \gamma_a^t)\, d\mu_{\mathbb{T}}(a), \qquad Mq \in L^1\left([s,t] \times \mathbb{T}^d\right),$$

where η_a^s, γ_a^t are the marginals of $\lambda_a^{s,t}$.

3.2 Notation and preliminary results

Measure-theoretic notation. We start by recalling some basic facts in Measure Theory. Let X, Y be *Polish* spaces, i.e. topological spaces whose topology is induced by a complete and separable distance. We endow a Polish space X with the corresponding Borel σ-algebra and denote by $\mathscr{P}(X)$ (resp. $\mathscr{M}_+(X)$, $\mathscr{M}(X)$) the family of Borel probability (resp. nonnegative and finite, real and with finite total variation) measures in X. For $A \subset X$ and $\mu \in \mathscr{M}(X)$ the *restriction* $\mu \llcorner A$ of μ to A is defined by $\mu \llcorner A(B) := \mu(A \cap B)$. We will denote by $\boldsymbol{i} : X \to X$ the identity map.

Definition 3.2.1 (Push-forward). *Let $\mu \in \mathscr{M}(X)$ and let $f : X \to Y$ be a Borel map. The push-forward $f_\# \mu$ is the measure in Y defined by $f_\# \mu(B) = \mu(f^{-1}(B))$ for any Borel set $B \subset Y$. The definition obviously extends, componentwise, to vector-valued measures.*

It is easy to check that $f_\# \mu$ has finite total variation as well, and that $|f_\# \mu| \le f_\# |\mu|$. An elementary approximation by simple functions shows the change of variable formula

$$\int_Y g \, d f_\# \mu = \int_X g \circ f \, d\mu \tag{3.2.1}$$

for any bounded Borel function (or even either nonnegative or nonpositive, and $\overline{\mathbb{R}}$-valued, in the case $\mu \in \mathscr{M}_+(X)$) $g : Y \to \mathbb{R}$.

Definition 3.2.2 (Narrow convergence and compactness). *Narrow (sequential) convergence in $\mathscr{P}(X)$ is the convergence induced by the duality with $C_b(X)$, the space of continuous and bounded functions in X. By Prokhorov theorem, a family $\mathscr{F}$ in $\mathscr{P}(X)$ is sequentially relatively compact with respect to the narrow convergence if and only if it is tight, i.e. for any $\varepsilon > 0$ there exists a compact set $K \subset X$ such that $\mu(X \setminus K) < \varepsilon$ for any $\mu \in \mathscr{F}$.*

In this chapter we use only the “easy” implication in Prokhorov theorem, namely that any tight family is sequentially relatively compact. It is immediate to check that a sufficient condition for tightness of a family $\mathscr{F}$ of probability measures is the existence of a *coercive* functional $\Psi : X \to [0, +\infty]$ (i.e. a functional such that its sublevel sets $\{\Psi \le t\}$, $t \in \mathbb{R}^+$, are relatively compact in X) such that

$$\int_X \Psi(x) d\mu(x) \le 1 \qquad \forall \mu \in \mathscr{F}.$$

Lemma 3.2.3 ([14], Lemma 2.4). *Let $\mu \in \mathscr{P}(X)$ and $\boldsymbol{u} \in L^2(X; \mathbb{R}^m)$. Then, for any Borel map $f : X \to Y$, $f_\#(\boldsymbol{u}\mu) \ll f_\# \mu$ and its density $\boldsymbol{v}$ with respect to $f_\# \mu$ satisfies*

$$\int_Y |\boldsymbol{v}|^2 \, d f_\# \mu \le \int_X |\boldsymbol{u}|^2 \, d\mu.$$

Furthermore, equality holds if and only if $\boldsymbol{u} = \boldsymbol{v} \circ f$ μ-a.e. in X.

Given $\mu \in \mathscr{M}_+(X\times Y)$, we shall denote by $\mu_x\otimes\lambda$ its *disintegration* via the projection map $\pi(x,y)=x$: here $\lambda=\pi_\#\mu\in\mathscr{M}_+(X)$, and $x\mapsto\mu_x\in\mathscr{P}(Y)$ is a Borel map (i.e. $x\mapsto\mu_x(A)$ is Borel for all Borel sets $A\subset Y$) characterized, up to λ-negligible sets, by

$$\int_{X\times Y} f(x,y)\,d\mu(x,y)=\int_X\left(\int_Y f(x,y)\,d\mu_x(y)\right)\,d\lambda(x) \tag{3.2.2}$$

for all nonnegative Borel map f. Conversely, any λ and any Borel map $x\mapsto\mu_x\in\mathscr{P}(Y)$ induce a probability measure μ in $X\times Y$ via (3.2.2).

Function spaces. We shall denote by $\Omega(D)$ the space $C([0,T];D)$, and by $\omega:[0,T]\to D$ its typical element. The evaluation maps at time t, $\omega\mapsto\omega(t)$, will be denoted by e_t.

If D is a smooth, compact Riemannian manifold without boundary (typically the d-dimensional flat torus $\mathbb{T}^d$), we shall denote μ_D its volume measure, and by d_D its Riemannian distance, normalizing the Riemannian metric so that μ_D is a probability measure. Although it does not fit exactly in this framework, we occasionally consider also the case $D=[0,1]^d$, because many results have already been obtained in this particular case.

We shall often consider measures $\boldsymbol{\eta}\in\mathscr{M}_+(\Omega(D))$ such that $(e_t)_\#\boldsymbol{\eta}\ll\mu_D$; in this case we shall denote by $\rho^{\boldsymbol{\eta}}:[0,T]\times D\to[0,+\infty]$ the density, characterized by

$$\rho^{\boldsymbol{\eta}}(t,\cdot)\mu_D:=(e_t)_\#\boldsymbol{\eta},\qquad t\in[0,T].$$

We denote by $\mathrm{SDiff}(D)$ the measure-preserving diffeomorphisms of D, and by $S(D)$ the measure-preserving maps in D:

$$S(D):=\{g:D\to D:\ g_\#\mu_D=\mu_D\}\,. \tag{3.2.3}$$

We also set

$$S^i(D):=\{g\in S(D):g\text{ is }\mu_D\text{-essentially injective}\}\,. \tag{3.2.4}$$

For any $g\in S^i(D)$ the inverse g^{-1} is well defined up to μ_D-negligible sets, μ_D-measurable, and $g^{-1}\circ g=\boldsymbol{i}=g\circ g^{-1}$ μ_D-a.e. in D. In particular, if $g\in S^i(D)$, $g^{-1}\in S^i(D)$.

We shall also denote by $\Gamma(D)$ the family of measure-preserving plans, i.e. the probability measures in $D\times D$ whose first and second marginal are μ_D:

$$\Gamma(D):=\{\gamma\in\mathscr{P}(D\times D):\ (\pi_1)_\#\gamma=\mu_D,\ (\pi_2)_\#\gamma=\mu_D\} \tag{3.2.5}$$

(here π_1, π_2 are the canonical coordinate projections).

Recall that $\mathrm{SDiff}(D)\subset S^i(D)\subset S(D)$ and that any element $g\in S(D)$ canonically induces a measure preserving plan γ_g, defined by

$$\gamma_g:=(\boldsymbol{i}\times g)_\#\mu_D.$$

Furthermore, this correspondence is continuous, as long as convergence in $L^2(\mu)$ of the maps g and narrow convergence of the plans are considered (see for instance Lemma 2.3 in [14]). Moreover

$$\overline{\{\gamma_g : \ g \in S^i(D)\}}^{\text{narrow}} = \Gamma(D), \tag{3.2.6}$$

$$\overline{\mathrm{SDiff}(D)}^{L^2(\mu_D)} = S(D) \qquad \text{if } D = [0,1]^d, \text{ with } d \geq 2 \tag{3.2.7}$$

(the first result is standard, see for example the explicit construction in [37, Theorem 1.4 (i)] in the case $D = [0,1]^d$, while the second one is proved in [37, Corollary 1.5])

The continuity equation. In the sequel we shall often consider weak solutions $\mu_t \in \mathscr{P}(D)$ of the continuity equation

$$\partial_t \mu_t + \mathrm{div}(\boldsymbol{v}_t \mu_t) = 0, \tag{3.2.8}$$

where $t \mapsto \mu_t$ is narrowly continuous (this is not restrictive, see for instance Lemma 8.1.2 of [11]) and $\boldsymbol{v}_t(x)$ is a suitable velocity field with $\|\boldsymbol{v}_t\|_{L^2(\mu_t)} \in L^1(0,T)$ (formally, $\boldsymbol{v}_t$ is a section of the tangent bundle and $|\boldsymbol{v}_t|$ is computed according to the Riemannian metric). The equation is understood in a weak (distributional) sense, by requiring that

$$\frac{d}{dt} \int_D \phi(t,x)\, d\mu_t(x) = \int_D \partial_t \phi + \langle \nabla \phi, \boldsymbol{v}_t \rangle \, d\mu_t \qquad \text{in } \mathcal{D}'(0,T)$$

for any $\phi \in C^1((0,T) \times D)$ with bounded first derivatives and support contained in $J \times D$, with $J \Subset (0,T)$. In the case when $D \subset \mathbb{R}^d$ is compact, we shall consider functions $\phi \in C^1\left((0,T) \times \mathbb{R}^d\right)$, again with support contained in $J \times \mathbb{R}^d$, with $J \Subset (0,T)$.

The following general principle allows to lift solutions of the continuity equation to measures in the space of continuous paths.

Theorem 3.2.4 (Superposition principle). *Assume that either D is a compact subset of $\mathbb{R}^d$, or D is a smooth compact Riemannian manifold without boundary, and let $\mu_t : [0,T] \to \mathscr{P}(D)$ be a narrowly continuous solution of the continuity equation* (3.2.8) *for a suitable velocity field $\boldsymbol{v}(t,x) = \boldsymbol{v}_t(x)$ satisfying $\|\boldsymbol{v}_t\|^2_{L^2(\mu_t)} \in L^1(0,T)$. Then there exists $\boldsymbol{\eta} \in \mathscr{P}(\Omega(D))$ such that*

(i) $\mu_t = (e_t)_\# \boldsymbol{\eta}$ for all $t \in [0,T]$;

(ii) the following energy inequality holds:

$$\int_{\Omega(D)} \int_0^T |\dot{\omega}(t)|^2 \, dt \, d\boldsymbol{\eta}(\omega) \leq \int_0^T \int_D |\boldsymbol{v}_t|^2 \, d\mu_t \, dt.$$

Proof. In the case when $D = \mathbb{R}^d$ (and therefore also when $D \subset \mathbb{R}^d$ is closed) this result is proved in Theorem 8.2.1 of [11] (see also [16], [123], [21] for related results). In the case when D is a smooth, compact Riemannian manifold we recover the same result thanks to an isometric embedding in $\mathbb{R}^m$, for m large enough. □

3.3 Variational models for generalized geodesics

3.3.1 Arnold's least action problem

Let f, $h \in \mathrm{SDiff}(D)$ be given. Following Arnold [15], we define $\delta^2(f,h)$ by minimizing the action

$$\mathscr{A}_T(g) := T\int_0^T \int_D \frac{1}{2}|\dot g(t,x)|^2\,d\mu_D(x)\,dt,$$

among all smooth curves

$$[0,T] \ni t \mapsto g(t,\cdot) \in \mathrm{SDiff}(D)$$

connecting f to h. By time rescaling, δ is independent of T. Since right composition with a given element $g \in \mathrm{SDiff}(D)$ does not change the action (as it amounts just to a relabelling of the initial position with g), the distance δ is right invariant, so it will be often useful to assume, in the minimization problem, that f is the identity map.

The action $\mathscr{A}_T$ can also be computed in terms of the velocity field $\boldsymbol{u}$, defined by $\boldsymbol{u}(t,x) = \dot g(t,y)|_{y=g^{-1}(t,x)}$, as

$$\mathscr{A}_T(\boldsymbol{u}) = T\int_0^T \int_D \frac{1}{2}|\boldsymbol{u}(t,x)|^2\,d\mu_D(x)\,dt.$$

As we mentioned in the introduction, connections between this minimization problem and (3.1.1) were achieved first by Ebin and Marsden, and then by Brenier: in [31], [35] he proved that if $(\boldsymbol{u},p)$ is a smooth solution of the Euler equation in $[0,T]\times D$, with $D=[0,1]^d$, and the inequality in (3.1.5) is strict, then the flow $g(t,x)$ of $\boldsymbol{u}$ is the unique solution of Arnold's minimization problem with $f=\boldsymbol{i}$, $h=g(T,\cdot)$.

By integrating the inequality $d_D^2(h(x),f(x)) \le \int_0^1 |\dot g(t,x)|^2\,dt$ one immediately obtains that $\|h-f\|_{L^2(D)} \le \sqrt{2}\delta(f,h)$; Shnirelman proved in [122] that in the case $D=[0,1]^d$ with $d\ge 3$ the Arnold distance is topologically equivalent to the L^2 distance: namely, there exist $C>0$, $\alpha>0$ such that

$$\delta(f,g) \le C\|f-g\|^\alpha_{L^2(D)} \qquad \forall f,\, g\in \mathrm{SDiff}(D). \tag{3.3.1}$$

Shnirelman also proved in [121] that when $d\ge 3$ the infimum is not attained in general and that, when $d=2$, $\delta(\boldsymbol{i},h)$ need not be finite (i.e., there exist $h\in\mathrm{SDiff}(D)$ which cannot be connected to $\boldsymbol{i}$ by a path with finite action).

3.3.2 Brenier's Lagrangian model and its extensions

In [31], Brenier proposed a relaxed version of the Arnold geodesic problem, and here we present more general versions of Brenier's relaxed problem, allowing first for final data in $\Gamma(D)$, and then for initial and final data in $\Gamma(D)$.

Let $\gamma \in \Gamma(D)$ be given; the class of admissible paths, called by Brenier *generalized incompressible flows*, is made by the probability measures $\boldsymbol{\eta}$ on $\Omega(D)$ such that

$$(e_t)_\# \boldsymbol{\eta} = \mu_D \qquad \forall t \in [0,T].$$

Then the action of an admissible $\boldsymbol{\eta}$ is defined as

$$\mathscr{A}_T(\boldsymbol{\eta}) := \int_{\Omega(D)} \mathscr{A}_T(\omega)\, d\boldsymbol{\eta}(\omega),$$

where

$$\mathscr{A}_T(\omega) := \begin{cases} T\int_0^T \frac{1}{2}|\dot\omega(t)|^2\, dt & \text{if } \omega \text{ is absolutely continuous in } [0,T] \\ +\infty & \text{otherwise,} \end{cases} \tag{3.3.2}$$

and $\overline{\delta}^2(\gamma_{\boldsymbol{i}}, \gamma)$ is defined by minimizing $\mathscr{A}_T(\boldsymbol{\eta})$ among all generalized incompressible flows $\boldsymbol{\eta}$ connecting $\gamma_{\boldsymbol{i}}$ to γ, i.e. those satisfying

$$(e_0, e_T)_\# \boldsymbol{\eta} = \gamma. \tag{3.3.3}$$

Notice that it is not clear, in this purely Lagrangian formulation, how the relaxed distance $\overline{\delta}(\eta,\gamma)$ between two measure preserving plans might be defined, not even when η and γ are induced by maps g, h. Only when $g \in S^i(D)$ we might use the right invariance and define $\overline{\delta}(\gamma_g, \gamma_h) := \overline{\delta}(\gamma_{\boldsymbol{i}}, \gamma_{h\circ g^{-1}})$.

These remarks led us to the following more general problem: let us denote

$$\tilde\Omega(D) := \Omega(D) \times D,$$

whose typical element will be denoted by (ω, a), and let us denote by $\pi_D : \tilde\Omega(D) \to D$ the canonical projection. We consider probability measures $\boldsymbol{\eta}$ in $\tilde\Omega(D)$ having μ_D as second marginal, i.e. $(\pi_D)_\#\boldsymbol{\eta} = \mu_D$; they can be canonically represented as $\boldsymbol{\eta}_a \otimes \mu_D$, where $\boldsymbol{\eta}_a \in \mathscr{P}(\Omega(D))$. The incompressibility constraint now becomes

$$\int_D (e_t)_\# \boldsymbol{\eta}_a \, d\mu_D(a) = \mu_D \qquad \forall t \in [0,T], \tag{3.3.4}$$

or equivalently $(e_t)_\#\boldsymbol{\eta} = \mu_D$ for all t, if we consider e_t as a map defined on $\tilde\Omega(D)$. Given initial and final data $\eta = \eta_a \otimes \mu_D$, $\gamma = \gamma_a \otimes \mu_D \in \Gamma(D)$, the constraint (3.3.3) now becomes

$$(e_0, \pi_D)_\#\boldsymbol{\eta} = \eta_a \otimes \mu_D, \qquad (e_T, \pi_D)_\#\boldsymbol{\eta} = \gamma_a \otimes \mu_D. \tag{3.3.5}$$

Equivalently, in terms of $\boldsymbol{\eta}_a$ we can write

$$(e_0)_\#\boldsymbol{\eta}_a = \eta_a, \qquad (e_T)_\#\boldsymbol{\eta}_a = \gamma_a. \tag{3.3.6}$$

Then, we define $\overline{\delta}^2(\eta,\gamma)$ by minimizing the action

$$\int_{\tilde\Omega(D)} \mathscr{A}_T(\omega)\,d\boldsymbol{\eta}(\omega,a)$$

among all generalized incompressible flows $\boldsymbol{\eta}$ (according to (3.3.4)) connecting η to γ (according to (3.3.5) or (3.3.6)). Notice that $\overline{\delta}^2$ is independent of T, because the action is scaling invariant; so we can use any interval $[a,b]$ in place of $[0,T]$ to define $\overline{\delta}$, and in this case we shall talk of generalized flow between η and γ in $[a,b]$ (this extension will play a role in Remark 3.3.2 below).

When $\eta_a=\delta_a$ (i.e. $\eta=\gamma_{\boldsymbol{i}}$), (3.3.6) tells us that almost all trajectories of $\boldsymbol{\eta}_a$ start from a: then $\int_D \boldsymbol{\eta}_a\,d\mu_D(a)$ provides us with a solution of Brenier's original model with the same action, connecting $\gamma_{\boldsymbol{i}}$ to γ. Conversely, any solution $\boldsymbol{\nu}$ of this model can be written as $\int_D \boldsymbol{\nu}_a\,d\mu_D$, with $\boldsymbol{\nu}_a$ concentrated on the curves starting at a, and $\boldsymbol{\nu}_a\otimes\mu_D$ provides us with an admissible path for our generalized problem, connecting $\gamma_{\boldsymbol{i}}$ to γ, with the same action.

Let us now analyze the properties of $(\Gamma(D),\overline{\delta})$; the fact that this is a metric space and even a *length space* (i.e. any two points can be joined by a geodesic with length equal to the distance) follows by the basic operations *reparameterization*, *restriction* and *concatenation* of generalized flows, that we are now going to describe.

Remark 3.3.1 (Repameterization). Let $\chi:[0,T]\to[0,T]$ be a C^1 map with $\dot\chi>0$, $\chi(0)=0$ and $\chi(T)=T$. Then, right composition of ω with χ induces a transformation $\boldsymbol{\eta}\mapsto\chi_\#\boldsymbol{\eta}$ between generalized incompressible flows that preserves the initial and final conditionl. As a consequence, if $\boldsymbol{\eta}$ is optimal the functional $\chi\mapsto\mathscr{A}_T(\chi_\#\boldsymbol{\eta})$ attains its minimum when $\chi(t)=t$. Changing variables we obtain

$$\mathscr{A}_T(\chi_\#\boldsymbol{\eta})=T\int_0^T\dot\chi^2(t)\int_{\tilde\Omega(D)}\frac12|\dot\omega|^2(\chi(t))\,d\boldsymbol{\eta}(\omega,a)\,dt=T\int_0^T\frac{1}{\dot g(s)}\int_{\tilde\Omega(D)}\frac12|\dot\omega|^2(s)\,d\boldsymbol{\eta}(\omega,a)\,ds$$

with $g=\chi^{-1}$. Therefore, choosing $g(s)=s+\varepsilon\phi(s)$, with $\phi\in C^1_c(0,T)$, the first variation gives

$$\int_0^T\left(\int_{\tilde\Omega(D)}|\dot\omega|^2(s)\,d\boldsymbol{\eta}(\omega,a)\right)\dot\phi(s)\,ds=0.$$

This proves that $s\mapsto\int_{\tilde\Omega(D)}|\dot\omega|^2(s)\,d\boldsymbol{\eta}(\omega,a)$ is equivalent to a constant. We shall call the square root of this quantity *speed* of $\boldsymbol{\eta}$.

Remark 3.3.2 (Restriction and concatenation). Let $[s,t]\subset[0,T]$ and let $r_{s,t}:C([0,T];D)\to C([s,t];D)$ be the restriction map. It is immediate to check that, for any generalized incompressible flow $\boldsymbol{\eta}=\boldsymbol{\eta}_a\otimes\mu_D$ in $[0,T]$ between η and γ, the measure

$(r_{s,t})_\#\boldsymbol{\eta}$ is a generalized incompressible flow in $[s,t]$ between $\eta_s := (e_s)_\#\boldsymbol{\eta}_a \otimes \mu_D$ and $\gamma_t := (e_t)_\#\boldsymbol{\eta}_a \otimes \mu_D$, with action equal to

$$(t-s)\int_{\tilde{\Omega}(D)}\int_s^t \frac{1}{2}|\dot{\omega}(\tau)|^2\,d\tau\,d\boldsymbol{\eta}(\omega,a).$$

Let $s < l < t$ and let $\boldsymbol{\eta} = \boldsymbol{\mu}_a \otimes \mu_D$, $\boldsymbol{\nu} = \boldsymbol{\nu}_a \otimes \mu_D$ be generalized incompressible flows, respectively defined in $[s,l]$ and $[l,t]$, and joining η to γ and γ to θ. Then, writing $\gamma_a = (e_l)_\#\boldsymbol{\eta}_a = (e_l)_\#\boldsymbol{\nu}_a$, we can disintegrate both $\boldsymbol{\eta}_a$ and $\boldsymbol{\nu}_a$ with respect to γ_a to obtain

$$\boldsymbol{\eta}_a = \int_D \boldsymbol{\eta}_{a,x}\,d\gamma_a(x) \in \mathscr{P}(C([s,l];D)), \qquad \boldsymbol{\nu}_a = \int_D \boldsymbol{\nu}_{a,x}\,d\gamma_a(x) \in \mathscr{P}(C([l,t];D)),$$

with $\boldsymbol{\eta}_{a,x}$, $\boldsymbol{\nu}_{a,x}$ concentrated on the curves ω with $\omega(l) = x$. We can then consider the image $\boldsymbol{\lambda}_{x,a}$, via the concatenation of paths (from the product of $C([s,l];D)$ and $C([l,t];D)$ to $C([s,t];D)$), of the product measure $\boldsymbol{\eta}_{a,x} \times \boldsymbol{\nu}_{a,x}$ to obtain a probability measure in $C([s,t];D)$ concentrated on paths passing through x at time l. Eventually, setting

$$\boldsymbol{\lambda} = \int_{D\times D} \boldsymbol{\lambda}_{x,a}\,d(\gamma_a \otimes \mu_D)(x,a),$$

we obtain a generalized incompressible flow in $[s,t]$ joining η to θ with action given by

$$\frac{t-s}{l-s}\mathscr{A}_{[s,l]}(\boldsymbol{\eta}) + \frac{t-s}{t-l}\mathscr{A}_{[l,t]}(\boldsymbol{\nu}),$$

where $\mathscr{A}_{[s,l]}(\boldsymbol{\eta})$ is the action of $\boldsymbol{\eta}$ in $[s,l]$ and $\mathscr{A}_{[l,t]}(\boldsymbol{\nu})$ is the action of $\boldsymbol{\nu}$ in $[l,t]$ (strictly speaking, the action of their restrictions).

A simple consequence of the previous remarks is that $\overline{\delta}$ is a distance in $\Gamma(D)$ (it suffices to concatenate flows with unit speed); in addition, the restriction of an optimal incompressible flow $\boldsymbol{\eta} = \boldsymbol{\eta}_a \otimes \mu_D$ between $\eta_a \otimes \mu_D$ and $\gamma_a \otimes \mu_D$ to an interval $[s,t]$ is still an optimal incompressible flow in $[s,t]$ between the plans $(e_s)_\#\boldsymbol{\eta}_a \otimes \mu_D$ and $(e_t)_\#\boldsymbol{\eta}_a \otimes \mu_D$. This property will be useful in Section 3.6.

Another important property of $\overline{\delta}$ that will be useful in Section 3.6 is its lower semicontinuity with respect to the narrow convergence, that we are going to prove in the next theorem. Another non-trivial fact is the existence of at least one generalized incompressible flow with finite action. In [31, Section 4] Brenier proved the existence of such a flow in the case $D = \mathbb{T}^d$. Then in [122, Section 2], using a (non-injective) Lipschitz measure-preserving map from $\mathbb{T}^d$ to $[0,1]^d$, Shnirelman produced a flow with finite action also in this case (see also [35, Section 3]). In the next theorem we will show how to construct a flow with finite action in a compact subset D whenever flows with finite action can be built in D' and a possibly non-injective, Lipschitz and measure-preserving map $f : D' \to D$ exists.

Theorem 3.3.3. *Assume that $D \subset \mathbb{R}^d$ is a compact set. Then the infimum in the definition of $\overline{\delta}(\eta,\gamma)$ is achieved,*

$$(\eta,\gamma) \mapsto \overline{\delta}(\eta,\gamma) \text{ is narrowly lower semicontinuous} \tag{3.3.7}$$

and

$$\overline{\delta}(\gamma_{\boldsymbol{i}},\gamma_h) \le \delta(\boldsymbol{i},h) \qquad \forall h \in \mathrm{SDiff}(D). \tag{3.3.8}$$

Furthermore, $\sup_{\eta,\,\gamma\in\Gamma(D)} \overline{\delta}(\eta,\gamma) \le \sqrt{d}$ *when either $D=[0,1]^d$ or $D=\mathbb{T}^d$ and, more generally,*

$$\sup_{\gamma\in\Gamma(D)} \overline{\delta}_D(\gamma_{\boldsymbol{i}},\gamma) \le \mathrm{Lip}(f) \sup_{\gamma'\in\Gamma(D')} \overline{\delta}_{D'}(\gamma_{\boldsymbol{i}},\gamma')$$

whenever a Lipschitz measure-preserving map $f: D' \to D$ exists.

Proof. The inequality $\overline{\delta}(\gamma_{\boldsymbol{i}},\gamma_h) \le \delta(\boldsymbol{i},h)$ simply follows by the fact that any smooth flow g induces a generalized one, with the same action, by the formula $\boldsymbol{\eta} = \Phi_\# \mu_D$, where $\Phi : D \to \tilde{\Omega}(D)$ is the map $x \mapsto (g(\cdot,x),x)$. Assuming that some generalized incompressible flow with a finite action between η and γ exists, the existence of an optimal one follows by the narrow lower semicontinuity of $\boldsymbol{\eta} \mapsto \mathscr{A}_T(\boldsymbol{\eta})$ (because $\omega \mapsto \mathscr{A}_T(\omega)$ is lower semicontinuous in $\Omega(D)$) and by the tightness of minimizing sequences (because $\mathscr{A}_T(\omega)$ is coercive in $\Omega(D)$, by the Ascoli-Arzelà theorem). A similar argument also proves the lower semicontinuity of $(\eta,\gamma) \mapsto \overline{\delta}(\eta,\gamma)$, as the conditions (3.3.4), (3.3.5) are stable under narrow convergence (of $\boldsymbol{\eta}$ and η, γ).

When either $D=[0,1]^d$ or $D=\mathbb{T}^d$, it follows by the explicit construction in [31], [122] that $\overline{\delta}(\gamma_{\boldsymbol{i}},\gamma_h) \le \sqrt{d}$ for all $h \in S(D)$; by right invariance (see Proposition 3.3.4 below) the same estimate holds for $\overline{\delta}(\gamma_f,\gamma_h)$ with $f \in S^i(D)$; by density and lower semicontinuity it extends to $\overline{\delta}(\eta,\gamma)$, with η, $\gamma \in \Gamma(D)$.

Let $f: D' \to D$ be a Lipschitz measure-preserving map and $h \in S(D)$; we claim that it suffices to show the existence of $\gamma' \in \Gamma(D')$ such that $(f\times f)_\#\gamma' = (\boldsymbol{i}\times h)_\#\mu_D$. Indeed, if this is proved, since f naturally induces by left composition a map F from $\tilde{\Omega}(D')$ to $\tilde{\Omega}(D)$ given by $(\omega(t),a) \mapsto (f(\omega(t)),a)$, then to any $\boldsymbol{\eta} \in \Omega(D')$ connecting $\boldsymbol{i}$ to γ' we can associate $F_\#\boldsymbol{\eta}$, which will be a generalized incompressible flow connecting $\boldsymbol{i}$ to h. By the trivial estimate

$$\mathscr{A}_T(F_\#\boldsymbol{\eta}) \le \mathrm{Lip}^2(f)\mathscr{A}_T(\boldsymbol{\eta}),$$

one obtains $\overline{\delta}_D(\gamma_{\boldsymbol{i}},h) \le \mathrm{Lip}(f)\overline{\delta}_{D'}(\gamma_{\boldsymbol{i}},\gamma')$. By density and lower semicontinuity we get the estimate on $\overline{\delta}_D(\gamma_{\boldsymbol{i}},\gamma)$ for all $\gamma \in \Gamma(D)$.

Thus, to conclude the proof, we have to construct γ'. Let us consider the disintegration of $\mu_{D'}$ induced by the map f, that is

$$\mu_{D'} = \int_D \mu_y \, d\mu_D(y) \tag{3.3.9}$$

where, for μ_D-a.e. y, μ_y is a probability measure in D' concentrated on the compact set $f^{-1}(y)$. We now define γ' as

$$\gamma' := \int_D \mu_y \times \mu_{h(y)}\, d\mu_D(y).$$

Clearly the first marginal of γ' is $\mu_{D'}$; since $h \in S(D)$, changing variables in (3.3.9) one has $\mu_{D'} = \int_D \mu_{h(y)}\, d\mu_D(y)$, and so also the second marginal of γ' is μ_D. Let us now prove that $(f \times f)_\# \gamma' = (\boldsymbol{i} \times h)_\# \mu_D$: for any $\phi \in C_b(D \times D)$ we have

$$\begin{aligned}\int_{D\times D} \phi(y,y')\, d(f\times f)_\#\gamma'(y,y') &= \int_{D'\times D'} \phi(f(x),f(x'))\, d\gamma'(x,x')\\ &= \int_D \int_{D'\times D'} \phi(f(x),f(x'))\, d\mu_y(x)\, d\mu_{h(y)}(x')\, d\mu_D(y)\\ &= \int_D \phi(y,h(y))\, d\mu_D(y),\end{aligned}$$

where in the last equality we used that μ_y is concentrated on $f^{-1}(y)$ and $\mu_{h(y)}$ is concentrated on $f^{-1}(h(y))$ for μ_D-a.e. y. □

By (3.3.1), (3.3.8) and the narrow lower semicontinuity of $\overline{\delta}(\boldsymbol{i},\cdot)$ we get

$$\overline{\delta}(\gamma_{\boldsymbol{i}}, h) \le C\|h - \boldsymbol{i}\|^\alpha_{L^2(D)} \qquad \text{if } h \in S(D),\ D = [0,1]^d,\ d \ge 3. \tag{3.3.10}$$

We conclude this section by pointing out some additional properties of the metric space $(\Gamma(D), \overline{\delta})$.

Proposition 3.3.4. $(\Gamma(D), \overline{\delta})$ *is a complete metric space, whose convergence implies narrow convergence. Furthermore, the distance* $\overline{\delta}$ *is right invariant under the action of* $S^i(D)$ *on* $\Gamma(D)$*. Finally,* $\overline{\delta}$*-convergence is strictly stronger than narrow convergence and, as a consequence,* $(\Gamma(D), \overline{\delta})$ *is not compact.*

Proof. We will prove that $\overline{\delta}(\eta,\gamma) \ge W_2(\eta,\gamma)$, where W_2 is the quadratic Wasserstein distance in $\mathscr{P}(D\times D)$ (with the quadratic cost $c((x_1,x_2),(y_1,y_2)) = d_D^2(x_1,y_1)/2 + d_D^2(x_2,y_2)/2$); as this distance metrizes the narrow convergence, this will give the implication between $\overline{\delta}$-convergence and narrow convergence. In order to show the inequality $\overline{\delta}(\eta,\gamma) \ge W_2(\eta,\gamma)$ we consider an optimal flow $\boldsymbol{\eta}_a \otimes \mu_D$ defined in $[0,1]$; then, denoting by $\omega_a \in \Omega(D)$ the constant path identically equal to a, and by $\boldsymbol{\nu}_a \in \mathscr{P}(C([0,1];D\times D))$ the measure $\boldsymbol{\eta}_a \times \delta_{\omega_a}$, the measure $\boldsymbol{\nu} := \int_D \boldsymbol{\nu}_a\, d\mu_D(a) \in \mathscr{P}(C([0,1];D\times D))$ provides a "dynamical transference plan" connecting η to γ (i.e. $(e_0)_\#\boldsymbol{\nu} = \eta$, $(e_1)_\#\boldsymbol{\nu} = \gamma$, see [133, Chapter 7]) whose action is $\overline{\delta}^2(\eta,\gamma)$; since the action of any dynamical transference plan bounds from above $W_2^2(\eta,\gamma)$, the inequality is achieved.

The completeness of $(\Gamma(D), \overline{\delta})$ is a consequence of the inequality $\overline{\delta} \geq W_2$ (so that Cauchy sequences in this space are Cauchy sequences for the Wasserstein distance), the completeness of the Wasserstein spaces of probability measures and the narrow lower semicontinuity of $\overline{\delta}$: we leave the details of the simple proof to the reader.

The right invariance of $\overline{\delta}$ simply follows by the fact that $\eta \circ h = \eta_{h(a)} \otimes \mu_D$, $\gamma \circ h = \gamma_{h(a)} \otimes \mu_D$, so that

$$\overline{\delta}(\eta \circ h, \gamma \circ h) \leq \overline{\delta}(\eta, \gamma),$$

because we can apply the same transformation to any admissible flow $\boldsymbol{\eta}_a \otimes \mu_D$ connecting η to γ, producing an admissible flow $\boldsymbol{\eta}_{h(a)} \otimes \mu_D$ between $\eta \circ h$ and $\gamma \circ h$ with the same action. If $h \in S^i(D)$ the inequality can be reversed, using h^{-1}.

Now, let us prove the last part of the statement. We first show that

$$\frac{1}{2}\int_D d_D^2(f, h)\, d\mu_D \leq \overline{\delta}^2(\gamma_f, \gamma_h) \qquad \forall f,\ h \in S(D). \tag{3.3.11}$$

Indeed, considering again an optimal flow $\boldsymbol{\eta}_a \otimes \mu_D$, for μ_D-a.e. $a \in D$ we have

$$\frac{1}{2}d_D^2(f(a), h(a)) = W_2^2(\delta_{f(a)}, \delta_{h(a)}) \leq T \int_{\Omega(D)} \int_0^T \frac{1}{2}|\dot{\omega}(t)|^2\, dt\, d\boldsymbol{\eta}_a(\omega),$$

and we need only to integrate this inequality with respect to a. From (3.3.11) we obtain that $S(D)$ is a closed subset of $\Gamma(D)$, relative to the distance $\overline{\delta}$. In particular, considering for instance a sequence $(g_n) \subset S(D)$ narrowly converging to $\gamma \in \Gamma(D) \setminus S(D)$, whose existence is ensured by (3.2.6), one proves that the two topologies are not equivalent and the space is not compact. □

Combining right invariance with (3.3.10), we obtain

$$\overline{\delta}(\gamma_g, \gamma_h) = \overline{\delta}(\gamma_{\boldsymbol{i}}, \gamma_{h\circ g^{-1}}) \leq C\|g - h\|^{\alpha}_{L^2(D)} \qquad \forall h \in S(D),\ g \in S^i(D) \tag{3.3.12}$$

if $D = [0, 1]^d$ with $d \geq 3$. By the density of $S^i(D)$ in $S(D)$ in the L^2 norm and the lower semicontinuity of $\overline{\delta}$, this inequality still holds when $g \in S(D)$.

3.3.3 Brenier's Eulerian-Lagrangian model

In [35], Brenier proposed a second possible relaxation of Arnold's problem, motivated by the fact that this second relaxation allows for a much more precise description of the pressure field, compared to the Lagrangian model (see Section 3.6).

Still denoting by $\eta = \eta_a \otimes \mu_D \in \Gamma(D)$, $\gamma = \gamma_a \otimes \mu_D \in \Gamma(D)$ the initial and final plan, respectively, the idea is to add to the Eulerian variable x a Lagrangian one a (which, in the case $\eta = \gamma_{\boldsymbol{i}}$, simply labels the position of the particle at time 0) and to consider the family of distributional solutions, indexed by $a \in D$, of the continuity equation

$$\partial_t c_{t,a} + \mathrm{div}(\boldsymbol{v}_{t,a} c_{t,a}) = 0 \qquad \text{in } \mathcal{D}'((0,T) \times D), \quad \text{for } \mu_D\text{-a.e. } a, \tag{3.3.13}$$

with the initial and final conditions

$$c_{0,a} = \eta_a, \qquad c_{T,a} = \gamma_a, \qquad \text{for } \mu_D\text{-a.e. } a. \tag{3.3.14}$$

Notice that minimization of the kinetic energy $\int_0^T \int_D |\boldsymbol{v}_{t,a}|^2 \, dc_{t,a} \, dt$ among all possible solutions of the continuity equation would give, according to [19], the optimal transport problem between η_a and γ_a (for instance, a path of Dirac masses on a geodesic connecting $g(a)$ to $h(a)$ if $\eta_a = \delta_{g(a)}$, $\gamma_a = \delta_{h(a)}$). Here, instead, by averaging with respect to a we minimize the *mean* kinetic energy

$$\int_D \int_0^T \int_D |\boldsymbol{v}_{t,a}|^2 \, dc_{t,a} \, dt \, d\mu_D(a)$$

with the only *global* constraint between the family $\{c_{t,a}\}$ given by the incompressibility of the flow:

$$\int_D c_{t,a} \, d\mu_D(a) = \mu_D \qquad \forall t \in [0,T]. \tag{3.3.15}$$

It is useful to rewrite this minimization problem in terms of the the global measure c in $[0,T] \times D \times D$ and the measures c_t in $D \times D$

$$c := c_{t,a} \otimes (\mathscr{L}^1 \times \mu_D), \qquad c_t := c_{t,a} \otimes \mu_D$$

(from whom $c_{t,a}$ can obviously be recovered by disintegration), and the velocity field $\boldsymbol{v}(t,x,a) := \boldsymbol{v}_{t,a}(x)$: the action becomes

$$\mathscr{A}_T(c, \boldsymbol{v}) := T \int_0^T \int_{D \times D} \frac{1}{2} |\boldsymbol{v}(t,x,a)|^2 \, dc(t,x,a),$$

while (3.3.13) is easily seen to be equivalent to

$$\frac{d}{dt} \int_{D \times D} \phi(x,a) \, dc_t(x,a) = \int_{D \times D} \langle \nabla_x \phi(x,a), \boldsymbol{v}(t,x,a) \rangle \, dc_t(x,a) \tag{3.3.16}$$

for all $\phi \in C_b(D \times D)$ with a bounded gradient with respect to the x variable.

Thus, we can minimize the action on the class of couples measures-velocity fields $(c, \boldsymbol{v})$ that satisfy (3.3.16) and (3.3.15), with the endpoint condition (3.3.14). The existence of a minimum in this class can be proved by standard compactness and lower semicontinuity arguments (see [35] for details). This minimization problem leads to a squared distance between η and γ, that we shall still denote by $\overline{\delta}^2(\eta, \gamma)$. Our notation is justified by the essential equivalence of the two models, proved in the next section.

3.4 Equivalence of the two relaxed models

In this section we show that the Lagrangian model is equivalent to the Eulerian-Lagrangian one, in the sense that minimal values are the same, and there is a way (not canonical, in one direction) to pass from minimizers of one problem to minimizers of the other one.

Theorem 3.4.1. *With the notations of Sections 3.3.2 and 3.3.3,*

$$\min_{\boldsymbol{\eta}} \mathscr{A}_T(\boldsymbol{\eta}) = \min_{(c,\boldsymbol{v})} \mathscr{A}_T(c,\boldsymbol{v})$$

for any η, $\gamma \in \Gamma(D)$. More precisely, any minimizer $\boldsymbol{\eta}$ of the Lagrangian model connecting η to γ induces in a canonical way a minimizer $(c,\boldsymbol{v})$ of the Eulerian-Lagrangian one, and satisfies for $\mathscr{L}^1$-a.e. $t \in [0,T]$ the condition

$$\dot\omega(t) = \boldsymbol{v}_{t,a}(e_t(\omega)) \qquad \textit{for } \boldsymbol{\eta}\textit{-a.e. } (\omega,a). \tag{3.4.1}$$

Proof. Up to an isometric embedding, we shall assume that $D \subset \mathbb{R}^m$ isometrically (this is needed to apply Lemma 3.2.3). If $\boldsymbol{\eta} = \boldsymbol{\eta}_a \otimes \mu \in \mathscr{P}(\tilde\Omega(D))$ is a generalized incompressible flow, we denote by $D' \subset D$ a Borel set of full measure such that $\mathscr{A}_T(\boldsymbol{\eta}_a) < \infty$ for all $a \in D'$. For any $a \in D'$ we define

$$c^{\boldsymbol{\eta}}_{t,a} := (e_t)_\# \boldsymbol{\eta}_a, \qquad \boldsymbol{m}^{\boldsymbol{\eta}}_{t,a} = (e_t)_\# \left(\dot\omega(t)\boldsymbol{\eta}_a\right).$$

Notice that $\boldsymbol{m}^{\boldsymbol{\eta}}_{t,a}$ is well defined for $\mathscr{L}^1$-a.e. t, and absolutely continuous with respect to $c^{\boldsymbol{\eta}}_{t,a}$, thanks to Lemma 3.2.3; moreover, denoting by $\boldsymbol{v}^{\boldsymbol{\eta}}_{t,a}$ the density of $\boldsymbol{m}^{\boldsymbol{\eta}}_{t,a}$ with respect to $c^{\boldsymbol{\eta}}_{t,a}$, by the same lemma we have

$$\int_D |\boldsymbol{v}^{\boldsymbol{\eta}}_{t,a}|^2 \, dc^{\boldsymbol{\eta}}_{t,a} \le \int_{\Omega(D)} |\dot\omega(t)|^2 \, d\boldsymbol{\eta}_a(\omega), \tag{3.4.2}$$

with equality only if $\dot\omega(t) = \boldsymbol{v}^{\boldsymbol{\eta}}_{t,a}(e_t(\omega))$ for $\boldsymbol{\eta}_a$-a.e. ω. Then, we define the global measure and velocity by

$$c^{\boldsymbol{\eta}} := c^{\boldsymbol{\eta}}_{t,a} \otimes (\mathscr{L}^1 \times \mu_D), \qquad \boldsymbol{v}^{\boldsymbol{\eta}}(t,x,a) = \boldsymbol{v}^{\boldsymbol{\eta}}_t(x,a) := \boldsymbol{v}^{\boldsymbol{\eta}}_{t,a}(x).$$

It is easy to check that $(c^{\boldsymbol{\eta}}, \boldsymbol{v}^{\boldsymbol{\eta}})$ is admissible: indeed, writing $\eta = \eta_a \otimes \mu$, $\gamma = \gamma_a \otimes \mu_D$, the conditions $(e_0)_\# \boldsymbol{\eta}_a = \eta_a$ and $(e_T)_\# \boldsymbol{\eta}_a = \gamma_a$ yield $c^{\boldsymbol{\eta}}_{0,a} = \eta_a$ and $c^{\boldsymbol{\eta}}_{T,a} = \gamma_a$ (for μ_D-a.e. a).

This proves that (3.3.14) is fulfilled; the incompressibility constraint (3.3.15) simply comes from (3.3.4). Finally, we check (3.3.13) for $a \in D'$; this is equivalent, recalling the definition of $\boldsymbol{v}_{t,a}$, to

$$\frac{d}{dt} \int_D \phi(x) \, dc^{\boldsymbol{\eta}}_{t,a}(x) = \int_D \langle \nabla\phi, \boldsymbol{m}^{\boldsymbol{\eta}}_{t,a} \rangle, \tag{3.4.3}$$

which in turn corresponds to

$$\frac{d}{dt}\int_{\Omega(D)} \phi(\omega(t))\, d\boldsymbol{\eta}_a(\omega) = \int_{\Omega(D)} \langle \nabla\phi(\omega(t)), \dot\omega(t)\rangle\, d\boldsymbol{\eta}_a(\omega). \tag{3.4.4}$$

This last identity is a direct consequence of an exchange of differentiation and integral.

By integrating (3.4.2) in time and with respect to a we obtain that $\mathscr{A}_T(c^{\boldsymbol{\eta}}, \boldsymbol{v}^{\boldsymbol{\eta}}) \leq \mathscr{A}_T(\boldsymbol{\eta})$, and equality holds only if (3.4.1) holds.

So, in order to conclude the proof, it remains to find, given a couple measure-velocity field $(c, \boldsymbol{v})$ with finite action that satisfies (3.3.13), (3.3.14) and (3.3.15), an admissible generalized incompressible flow $\boldsymbol{\eta}$ with $\mathscr{A}_T(\boldsymbol{\eta}) \leq \mathscr{A}_T(c, \boldsymbol{v})$. By applying Theorem 3.2.4 to the family of solutions of the continuity equations (3.3.13), we obtain probability measures $\boldsymbol{\eta}_a$ with $(e_t)_\#\boldsymbol{\eta}_a = c_{t,a}$ and

$$\int_{\Omega(D)}\int_0^T |\dot\omega(t)|^2\, dt\, d\boldsymbol{\eta}_a(\omega) \leq \int_0^T \int_D |\boldsymbol{v}(t,x,a)|^2\, dc_{t,a}(x)\, dt. \tag{3.4.5}$$

Then, because of (3.3.15), it is easy to check that $\boldsymbol{\eta} := \boldsymbol{\eta}_a \otimes \mu_D$ is a generalized incompressible flow, and moreover $\boldsymbol{\eta}$ connects η to γ. By integrating (3.4.5) with respect to a, we obtain that $\mathscr{A}_T(\boldsymbol{\eta}) \leq \mathscr{A}_T(c, \boldsymbol{v})$. □

3.5 Comparison of metrics and gap phenomena

Throughout this section we shall assume that $D = [0,1]^d$. In [122], Shnirelman proved when $d \geq 3$ the following remarkable approximation theorem for Brenier's generalized (Lagrangian) flows:

Theorem 3.5.1. *If $d \geq 3$, then each generalized incompressible flow $\boldsymbol{\eta}$ connecting $\boldsymbol{i}$ to $h \in \mathrm{SDiff}(D)$ may be approximated together with the action by a sequence of smooth flows $(g_k(t,\cdot))$ connecting $\boldsymbol{i}$ to h. More precisely:*

(i) the measures $\boldsymbol{\eta}_k := (g_k(\cdot,x))_\#\mu_D$ narrowly converge in $\Omega(D)$ to $\boldsymbol{\eta}$;

(ii) $\mathscr{A}_T(g_k) = \mathscr{A}_T(\boldsymbol{\eta}_k) \to \mathscr{A}_T(\boldsymbol{\eta})$.

This result yields, as a byproduct, the identity

$$\overline{\delta}(\gamma_{\boldsymbol{i}}, \gamma_h) = \delta(\boldsymbol{i}, h) \qquad \text{for all } h \in \mathrm{SDiff}(D),\ d \geq 3. \tag{3.5.1}$$

More generally the relaxed distance $\overline{\delta}(\eta,\gamma)$ arising from the Lagrangian model can be compared, at least when $\eta=\gamma_{\boldsymbol{i}}$ and the final condition γ is induced by a map $h\in S(D)$, with the relaxation δ_* of the Arnold distance:

$$\delta_*(h) := \inf\left\{\liminf_{n\to\infty}\delta(\boldsymbol{i},h_n):\ h_n\in \mathrm{SDiff}(D),\ \int_D |h_n-h|^2\,d\mu_D\to 0\right\}. \tag{3.5.2}$$

By (3.3.7) and (3.3.8), we have $\delta_*(h)\geq \overline{\delta}(\gamma_{\boldsymbol{i}},\gamma_h)$, and a *gap phenomenon* is said to occur if the inequality is strict.

In the case $d=2$, while examples of $h\in \mathrm{SDiff}(D)$ such that $\delta(\boldsymbol{i},h)=+\infty$ are known [121], the nature of $\delta_*(h)$ and the possible occurrence of the gap phenomenon are not clear.

In this section we prove the non-occurrence of the gap phenomenon when the final condition belongs to $S(D)$, and even when it is a transport plan, still under the assumption $d\geq 3$. To this aim, we first extend the definition of δ_* by setting

$$\delta_*(\gamma) := \inf\left\{\liminf_{n\to\infty}\delta(\boldsymbol{i},h_n):\ h_n\in \mathrm{SDiff}(D),\ \gamma_{h_n}\to\gamma \text{ narrowly}\right\}. \tag{3.5.3}$$

This extends the previous definition (3.5.2), taking into account that γ_{h_n} narrowly converge to γ_h if and only if $h_n\to h$ in $L^2(\mu_D)$ (for instance, this is a simple consequence of [14, Lemma 2.3]).

Theorem 3.5.2. *If $d\geq 3$, then $\delta_*(\gamma)=\overline{\delta}(\gamma_{\boldsymbol{i}},\gamma)$ for all $\gamma\in\Gamma(D)$.*

The proof of the theorem, given at the end of this section, is a direct consequence of Theorem 3.5.1 and of the following approximation result of generalized incompressible flows by measure-preserving maps (possibly not smooth, or not injective), valid in *any* number of dimensions.

Theorem 3.5.3. *Let $\gamma\in\Gamma(D)$. Then, for any probability measure $\boldsymbol{\eta}$ on $\Omega(D)$ such that*

$$(e_t)_\#\boldsymbol{\eta}=\mu_D\quad \forall t\in[0,T],\qquad\qquad (e_0,e_T)_\#\boldsymbol{\eta}=\gamma,$$

and $\mathscr{A}_T(\boldsymbol{\eta})<\infty$, there exists a sequence of flows $(g_k(t,\cdot))_{k\in\mathbb{N}}\subset W^{1,2}\left([0,T];L^2(D)\right)$ such that:

(i) $g_k(t,\cdot)\in S(D)$ for all $t\in[0,T]$, hence $\boldsymbol{\eta}_k:=(\Phi_{g_k})_\#\mu_D$, with $\Phi_{g_k}(x)=g_k(\cdot,x)$, are generalized incompressible flows;

(ii) $\boldsymbol{\eta}_k$ narrowly converge in $\Omega(D)$ to $\boldsymbol{\eta}$ and $\mathscr{A}_T(g_k)=\mathscr{A}_T(\boldsymbol{\eta}_k)\to\mathscr{A}_T(\boldsymbol{\eta})$.

Proof. The first three steps of the proof are more or less the same as in the proof of Shnirelman's approximation theorem (Theorem 3.5.1 in [122]).

Step 1. Given $\varepsilon > 0$ small, consider the affine transformation of D into the concentric cube D_ε of size $1 - 4\varepsilon$:

$$T_\varepsilon(x) := (2\varepsilon, \dots, 2\varepsilon) + (1 - 4\varepsilon)x.$$

This transformation induces a map $\tilde{T}_\varepsilon$ from $\Omega(D)$ into $C([0,T]; D_\varepsilon)$ (which is indeed a bijection) given by

$$\tilde{T}_\varepsilon(\omega)(t) := T_\varepsilon(\omega(t)) \qquad \forall \omega \in \Omega(D).$$

Then we define $\tilde{\boldsymbol{\eta}}_\varepsilon := (\tilde{T}_\varepsilon)_\# \boldsymbol{\eta}$, and

$$\boldsymbol{\eta}_\varepsilon := (1 - 4\varepsilon)^d \tilde{\boldsymbol{\eta}}_\varepsilon + \boldsymbol{\eta}_{0,\varepsilon},$$

where $\boldsymbol{\eta}_{0,\varepsilon}$ is the "steady" flow in $D \setminus D_\varepsilon$: it consists of all the curves in $D \setminus D_\varepsilon$ that do not move for $0 \le t \le T$. It is then not difficult to prove that $\boldsymbol{\eta}_\varepsilon \to \boldsymbol{\eta}$ narrowly and $\mathscr{A}_T(\boldsymbol{\eta}_\varepsilon) \to \mathscr{A}_T(\boldsymbol{\eta})$, as $\varepsilon \to 0$.

Therefore, by a diagonal argument, it suffices to prove our theorem for a measure $\boldsymbol{\eta}$ which is steady near ∂D. More precisely we can assume that, if $\omega(0)$ is in the 2ε-neighborhood of ∂D, then $\omega(t) \equiv \omega(0)$ for $\boldsymbol{\eta}$-a.e. ω. Moreover, arguing as in Step 1 of the proof of the above mentioned approximation theorem in [122], we can assume that the flow does not move for $0 \le t \le \varepsilon$, that is, for $\boldsymbol{\eta}$-a.e. ω, $\omega(t) \equiv \omega(0)$ for $0 \le t \le \varepsilon$.

Step 2. Let us now consider a family of independent random variables $\omega_1, \omega_2, \dots$ defined in a common probability space $(Z, \mathcal{Z}, P)$, with values in $C([0,T], D)$ and having the same law $\boldsymbol{\eta}$. Recall that $\boldsymbol{\eta}$ is steady near ∂D and for $0 \le t \le \varepsilon$, so we can see ω_i as random variables with values in the subset of $\Omega(D)$ given by the curves which do not move for $0 \le t \le \varepsilon$ and in the 2ε-neighbourhood of the ∂D. By the law of large numbers, the random probability measures in $\Omega(D)$

$$\boldsymbol{\nu}_N(z) := \frac{1}{N} \sum_{i=1}^{N} \delta_{\omega_i(z)}, \qquad z \in Z,$$

narrowly converge to $\boldsymbol{\eta}$ with probability 1. Moreover, always by the law of large numbers, also

$$\mathscr{A}_T(\boldsymbol{\nu}_N(z)) \to \mathscr{A}_T(\boldsymbol{\eta})$$

with probability 1. Thus, choosing properly z, we have approximated $\boldsymbol{\eta}$ with measures $\boldsymbol{\nu}_N$ concentrated on a finite number of trajectories $\omega_i(z)(\cdot)$ which are steady in $[0, \varepsilon]$ and close to ∂D. From now on (as typical in Probability theory) the parameter z will be tacitly understood.

Step 3. Let $\varphi \in C_c^\infty(\mathbb{R}^d)$ be a smooth radial convolution kernel with $\varphi(x) = 0$ for $|x| \geq 1$ and $\varphi(x) > 0$ for $|x| < 1$. Given a finite number of trajectories $\omega_1, \ldots, \omega_N$ as described is step 2, we define

$$a_i(x) := \frac{1}{\varepsilon^d}\varphi\left(\frac{x - \omega_i(0)}{\varepsilon}\right) \qquad \text{if } \operatorname{dist}(\omega_i(0), \partial D) \geq \varepsilon,$$

$$a_i(x) := \frac{1}{\varepsilon^d}\sum_{\gamma\in\Gamma}\varphi\left(\frac{x - \gamma(\omega_i(0))}{\varepsilon}\right) \qquad \text{if } \operatorname{dist}(\omega_i(0), \partial D) \leq \varepsilon,$$

where Γ is the discrete group of motions in $\mathbb{R}^n$ generated by the reflections in the faces of D. It is easy to check that $\int a_i = 1$ and that $\operatorname{supp}(a_i)$ is the intersection of D with the closed ball $\overline{B}_\varepsilon(\omega_i(0))$. Define

$$g_{i,t}(x) := \omega_i(t) + (x - \omega_i(0)) \qquad \forall i = 1, \ldots, n.$$

Let $\mathscr{M}_N := (a_1, \ldots, a_N, g_{1,t}(x), \ldots, g_{N,t}(x))$ and let us consider the generalized flow $\boldsymbol{\eta}_N$ associated to $\mathscr{M}_N$, given by

$$\int_{\Omega(D)} f(\omega)\, d\boldsymbol{\eta}_N := \frac{1}{N}\sum_{i=1}^N \int_D a_i(x) f(t \mapsto g_{i,t}(x))\, dx \tag{3.5.4}$$

(that is, $\boldsymbol{\eta}_N$ is the measure in the space of paths given by $\frac{1}{N}\sum_i \int_D a_i(x)\delta_{g_{i,\cdot}(x)}dx$). The measure $\boldsymbol{\eta}_N$ is well defined for the following reason: if $\operatorname{dist}(\omega_i(0), \partial D) \leq \varepsilon$ we have $g_{i,t}(x) = x$, and if $\operatorname{dist}(\omega_i(0), \partial D) > \varepsilon$ and $a_i(x) > 0$ we still have that the curve $t \mapsto g_{i,t}(x)$ is contained in D because $a_i(x) > 0$ implies $|x - \omega_i(0)| \leq \varepsilon$ and, by construction, $\operatorname{dist}(\omega_i(t), \partial D) \geq \varepsilon$ for all times. Since the density $\rho^{\boldsymbol{\eta}_N}$ induced by $\boldsymbol{\eta}_N$ is given by

$$\rho^N(t, x) := \frac{1}{N}\sum_{i=1}^N a_i(x + \omega_i(0) - \omega_i(t)),$$

the flow $\boldsymbol{\eta}_N$ is not measure preserving. However we are more or less in the same situation as in Step 3 in the proof of the approximation theorem in [122] (the only difference being that we do not impose any final data). Thus, by [122, Lemma 1.2], with probability 1

$$\begin{aligned} &\sup_{x,t} |\rho^N(t, x) - 1| \to 0, \\ &\sup_{x,t} |\partial_x^\alpha \rho^N(t, x)| \to 0 \quad \forall \alpha, \\ &\int_D \int_0^T |\partial_t \rho^N(t, x)|^2\, dt\, dx \to 0 \end{aligned} \tag{3.5.5}$$

as $N \to \infty$. By the first two equations in (3.5.5), we can left compose $g_{i,t}$ with a smooth correcting flow $\zeta_t^N(x)$ as in Step 3 in the proof of the approximation theorem in [122], in such a way that the flow $\tilde{\boldsymbol{\eta}}_N$ associated to $\tilde{\mathscr{M}}_N := (a_1, \dots, a_N, \zeta_t^N \circ g_{1,t}(x), \dots, \zeta_t^N \circ g_{N,t}(x))$ via the formula analogous to (3.5.4) is incompressible. Moreover, thanks to the third equation in (3.5.5) and the convergence of $\mathscr{A}_T(\boldsymbol{\nu}_N)$ to $\mathscr{A}_T(\boldsymbol{\eta})$, one can prove that $\mathscr{A}_T(\tilde{\boldsymbol{\eta}}_N) \to \mathscr{A}_T(\boldsymbol{\eta})$ with probability 1.

We observe that, since $\boldsymbol{\eta}$ is steady for $0 \le t \le \varepsilon$, the same holds by construction for $\tilde{\boldsymbol{\eta}}_N$. Without loss of generality, we can therefore assume that ζ_t^N does not depend on t for $t \in [0, \varepsilon]$.

Step 4. In order to conclude, we see that the only problem now is that the flow $\tilde{\boldsymbol{\eta}}_N$ associated to $\tilde{\mathscr{M}}_N$ is still non-deterministic, since if $x \in \operatorname{supp}(a_i) \cap \operatorname{supp}(a_j)$ for $i \neq j$, then more that one curve starts from x. Let us partition D in the following way:

$$D = D_1 \cup D_2 \cup \ldots \cup D_L \cup E,$$

where E is $\mathscr{L}^d$-negligible, any set D_j is open, and all $x \in D_j$ belong to the interior of the supports of exactly $M = M(j) \le N$ sets a_i, indexed by $1 \le i_1 < \cdots < i_M \le N$ (therefore $L \le 2^N$). This decomposition is possible, as E is contained in the union of the boundaries of $\operatorname{supp} a_i$, which is $\mathscr{L}^d$-negligible.

Fix one of the sets D_j and assume just for notational simplicity that $i_k = k$ for $1 \le k \le M$. We are going to modify the flow $\tilde{\boldsymbol{\eta}}_N$ in D_j, increasing a little bit its action (say, by an amount $\alpha > 0$), in such a way that for each point in D_j only one curve starts from it. Given $x \in D_j$, we know that M curves start from it, weighted with mass $a_k(x) > 0$, and $\sum_{k=1}^M a_k(x) = 1$. These curves coincide for $0 \le t \le \varepsilon$ (since nothing moves), and then separate. We want to partition D_j in M sets E_k, with

$$\mathscr{L}^d(E_k) = \int_{D_j} a_k(x)\, dx, \qquad 1 \le k \le M$$

in such a way that, for any $x \in E_k$, only one curve ω_x^k starts from it at time 0, $\omega_x^k(t) \in D_j$ for $0 \le t \le \varepsilon$, and the map $E_k \ni x \mapsto \omega_x^k(\varepsilon) \in D_j$ pushes forward $\mathscr{L}^d \llcorner E_k$ into $a_k \mathscr{L}^d \llcorner D_j$. Moreover, we want the incompressibility condition to be preserved for all $t \in [0, \varepsilon]$. If this is possible, the proof will be concluded by gluing ω_x^k with the only curve starting from $\omega_x^k(\varepsilon)$ with weight $a_k(\omega^k(\varepsilon))$.

The above construction can be achieved in the following way. First we write the interior of D_j, up to null measure sets, as a countable union of disjoints open cubes (C_i) with size δ_i satisfying

$$\frac{M^2}{\varepsilon} \sum_i \frac{\delta_i^2}{\overline{b}_i^2} \mathscr{L}^d(C_i) \le \alpha, \tag{3.5.6}$$

with $\bar b_i := \min_{1\le k\le M}\min_{C_i} a_k$. This is done just considering the union of the grids in $\mathbb{R}^d$ given by $\mathbb{Z}^d/2^n$ for $n\in\mathbb{N}$, and taking initially our cubes in this family; if (3.5.6) does not hold, we keep splitting the cubes until it is satisfied ($\bar b_i$ can only increase under this additional splitting, therefore a factor 4 is gained in each splitting). Once this partition is given, the idea is to move the mass within each C_i for $0\le t\le\varepsilon$. At least heuristically, one can imagine that in C_i the functions a_k are almost constant and that the velocity of a generic path in C_i is at most of order δ_i/ε. Thus, the total energy of the new incompressible fluid in the interval $[0,\varepsilon]$ will be of order

$$\sum_i\int_{C_i}\int_0^\varepsilon|\dot\omega_x(t)|^2\,dt\,dx\le\frac{C}{\varepsilon}\sum_i\delta_i^2\mathscr{L}^d(C_i)$$

and the conclusion will follow by our choice of δ_i.

So, in order to make this argument rigorous, let us fix i and let us see how to construct our modified flow in the cube C_i for $t\in[0,\varepsilon]$. Slicing C_i with respect to the first $(d-1)$-variables, we see that the transport problem can be solved in each slice. Specifically, if C_i is of the form $x^i+(0,\delta_i)^d$, and we define

$$m^k:=\int_{C_i}a_k(x)\,dx,\qquad k=1,\dots,M,$$

whose sum is δ_i^d, then the points which belong to $C_i^k:=x^i+(0,\delta_i)^{d-1}\times J_k$ have to move along curves in order to push forward $\mathscr{L}^d\llcorner C_i^k$ into $a_k\mathscr{L}^d\llcorner C_i$, where J_k are M consecutive open intervals in $(0,\delta_i)$ with length $\delta_i^{1-d}m^k$. Moreover, this has to be done preserving the incompressibility condition.

If we write $x=(x',x_d)\in\mathbb{R}^d$ with $x'=(x_1,\dots,x_{d-1})$, we can transport the M uniform densities

$$\mathscr{H}^1\llcorner\big(x^i+\{x'\}\times J_k\big)\quad\text{with } x'\in[0,\delta_i]^{(d-1)},$$

into the M densities

$$a_k(x',\cdot)\mathscr{H}^1\llcorner\big(x^i+\{x'\}\times[0,\delta_i]\big)$$

moving the curves only in the d-th direction, i.e. keeping x' fixed. Thanks to Lemma 3.5.4 below and a scaling argument, we can do this construction paying at most $M^2\bar b_i^{-2}\delta_i^3/\varepsilon$ in each slice of C_i, and therefore with a total cost less than

$$\frac{M^2}{\varepsilon}\sum_i\frac{\delta_i^{d+2}}{\bar b_i^2}\le\alpha.$$

This concludes our construction. □

Lemma 3.5.4. *Let $M \geq 1$ be an integer and let $b_1, \ldots, b_M : [0,1] \to (0,1]$ be continuous with $\sum_1^M b_k = 1$. Setting $l_k = \int_0^1 b_k \, dt \in (0,1]$, and denoting by $J_1, \ldots, J_M$ consecutive intervals of $(0,1)$ with length l_k, there exists a family of uniformly Lipschitz maps $h(\cdot, x)$, with $h(t, \cdot) \in S([0,1])$, such that*

$$h(1, \cdot)_\#(\chi_{J_k} \mathscr{L}^1) = b_k \mathscr{L}^1, \qquad k = 1, \ldots, M$$

and

$$\mathscr{A}_1(h) \leq \frac{M^2}{\bar{b}^2}, \qquad \text{with } \bar{b} := \min_{1 \leq k \leq M} \min_{[0,1]} b_k > 0. \tag{3.5.7}$$

Proof. We start with a preliminary remark: let $J \subset (0,1)$ be an interval with length l and assume that $t \mapsto \rho_t$ is a nonnegative Lipschitz map between $[0,1]$ and $L^1(0,1)$, with $\rho_t \leq 1$ and $\int_0^1 \rho_t \, dx = l$ for all $t \in [0,1]$, and let $f(t, \cdot)$ be the unique (on J, up to countable sets) nondecreasing map pushing $\chi_J \mathscr{L}^1$ to ρ_t. Assume also that $\operatorname{supp} \rho_t$ is an interval and $\rho_t \geq r$ $\mathscr{L}^1$-a.e. on $\operatorname{supp} \rho_t$, with $r > 0$. Under this extra assumption, $f(t, x)$ is uniquely determined for all $x \in J$, and implicitly characterized by the conditions

$$\int_0^{f(t,x)} \rho_t(y) \, dy = \mathscr{L}^1((0, x) \cap J), \qquad f(t, x) \in \operatorname{supp} \rho_t.$$

This implies, in particular, that $f(\cdot, x)$ is continuous *for all* $x \in J$. We are going to prove that this map is even Lipschitz continuous in $[0,1]$ and

$$\left|\frac{d}{dt} f(t, x)\right| \leq \frac{\operatorname{Lip}(\rho_\cdot)}{r} \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in [0,1] \tag{3.5.8}$$

for all $x \in J$. To prove this fact, we first notice that the endpoints of the interval $\operatorname{supp} \rho_t$ (whose length is at least l) move at most with velocity $\operatorname{Lip}(\rho_\cdot)/r$; then, we fix $x \in J = [a, b]$ and consider separately the cases

$$x \in \partial J = \{a, b\}, \qquad x \in \operatorname{Int}(J) = (a, b).$$

In the first case, since for any $t \in [0,1]$

$$\int_0^{f(t,a)} \rho_t(y) \, dy = 0, \qquad \int_0^{f(t,b)} \rho_t(y) \, dy = \mathscr{L}^1(J),$$

and by assumption $f(t, x) \in \operatorname{supp} \rho_t$ for any $x \in J$, we get $\operatorname{supp} \rho_t = [f(t, a), f(t, b)]$ for all $t \in [0,1]$. This, together with the fact that the endpoints of the interval $\operatorname{supp} \rho_t$ move at most with velocity $\operatorname{Lip}(\rho_\cdot)/r$, implies (3.5.8) if $x \in \partial J$. In the second case we have

$$\int_0^{f(t,x)} \rho_t(y) \, dy \in (0, \mathscr{L}^1(J)),$$

therefore $f(t,x) \in \mathrm{Int}(\operatorname{supp}\rho_t)$ for all $t \in [0,1]$. It suffices now to find a Lipschitz estimate of $|f(s,x)-f(t,x)|$ when s, t are sufficiently close. Assume that $f(s,x) \le f(t,x)$: adding and subtracting $\int_0^{f(s,x)} \rho_t(y)\,dy$ in the identity

$$\int_0^{f(t,x)} \rho_t(y)\,dy = \int_0^{f(s,x)} \rho_s(y)\,dy$$

we obtain

$$\int_{f(s,x)}^{f(t,x)} \rho_t(y)\,dy = \int_0^{f(s,x)} \rho_s(y) - \rho_t(y)\,dy.$$

Now, as $f(s,x)$ belongs to $\operatorname{supp}\rho_t$ for $|s-t|$ sufficiently small, we get

$$r|f(s,x) - f(t,x)| \le \mathrm{Lip}(\rho_\cdot)|t-s|.$$

This proves the Lipschitz continuity of $f(\cdot,x)$ and (3.5.8).

Given this observation, to prove the lemma it suffices to find maps $t \mapsto \rho_t^k$ connecting $\chi_{J_k}\mathscr{L}^1$ to $b_k\mathscr{L}^1$ satisfying:

(i) $\operatorname{supp}\rho_t^k$ is an interval, and $\rho_t^k \ge \min_{[0,1]} b_k \ge \bar b$ $\mathscr{L}^1$-a.e. on its support;

(ii) $\mathrm{Lip}(\rho_\cdot^k) \le \frac{M-1}{2}$ on $[0,\frac12]$, and $\mathrm{Lip}(\rho_\cdot^k) \le 2$ on $[\frac12,1]$;

(iii) $\sum_{k=1}^{M} \rho_t^k = 1$ for all $t \in [0,1]$.

Indeed, this would produce maps with time derivative bounded by $(M-1)/(2\bar b)$ on $[0,\frac12]$ and bounded by $2/\bar b$ on $[\frac12,1]$, and this easily gives (3.5.7).

The construction can be achieved in two steps. First, we connect $\chi_{J_k}\mathscr{L}^1$ to $l_k\mathscr{L}^1$ in the time interval $[0,\frac12]$; then, we connect $l_k\mathscr{L}^1$ to $b_k\mathscr{L}^1$ in $[\frac12,1]$ by a linear interpolation. The Lipschitz constants of the second step are easily seen to be less than 2, so let us focus on the first interpolation.

Let us first consider the case of two densities $\rho^1 = \chi_{J_1}$ and $\rho^2 = \chi_{J_2}$, with $J_1 = (0,l_1)$ and $J_2 = (l_1,l)$. In the time interval $[0,\tau]$, we define the expanding intervals

$$J_{1,t} = (0, l_1 + \frac{t}{\tau} l_2), \quad J_{2,t} = (l_1 - \frac{t}{\tau} l_1, 1),$$

so that $J_{k,\tau} = (0,l)$ for $k = 1$, 2, and then define

$$\rho_t^1 := \begin{cases} 1 & \text{on } (0, l_1 - \frac{t}{\tau} l_1), \\ l_1/l & \text{on } (l_1 - \frac{t}{\tau} l_1, l_1 + \frac{t}{\tau} l_2), \\ 0 & \text{otherwise.} \end{cases} \qquad \rho_t^2 := \begin{cases} 1 & \text{on } (l_1 + \frac{t}{\tau} l_2, l), \\ l_2/l & \text{on } (l_1 - \frac{t}{\tau} l_1, l_1 + \frac{t}{\tau} l_2), \\ 0 & \text{otherwise.} \end{cases}$$

By construction $\rho_t^k \geq l_k$ on $J_{k,t}$ for $k = 1,\, 2$, $\rho_t^1 + \rho_t^2 = 1$, and it is easy to see that

$$\operatorname{Lip}(\rho_\cdot^k) \leq \frac{l_1 l_2}{\tau l} \leq \frac{l}{4\tau}. \tag{3.5.9}$$

We can now define the desired interpolation on $[0, \frac{1}{2}]$ for general $M \geq 2$. Let us define

$$t_i := \frac{i}{2(M-1)} \qquad \text{for } i = 1, \dots, M-1,$$

so that $t_{M-1} = \frac{1}{2}$. We will achieve our construction of ρ_t^k on $[0, \frac{1}{2}]$ in $M-1$ steps, where at each step we will progressively define ρ_t^k on the time interval $[t_{i-1}, t_i]$.

First, in the time interval $[0, t_1]$, we leave fixed $\rho_0^k := \chi_{J_k}\mathscr{L}^1$ for $k \geq 3$ (if such k exist), while we apply the above construction in $J_1 \cup J_2$ to ρ^1 and ρ^2. In this way, on $[0, t_1]$, $\rho_0^1 := \chi_{J_1}\mathscr{L}^1$ is connected to $\rho_{t_1}^1 := \frac{l_1}{l_1+l_2}\chi_{J_1\cup J_2}\mathscr{L}^1$, and $\rho_0^2 := \chi_{J_2}\mathscr{L}^1$ is connected to $\rho_{t_1}^2 := \frac{l_2}{l_1+l_2}\chi_{J_1\cup J_2}\mathscr{L}^1$.

Now, as a second step, we want to connect $\rho_{t_1}^k$ to $\frac{l_k}{l_1+l_2+l_3}\chi_{J_1\cup J_2\cup J_3}\mathscr{L}^1$ for $k = 1,\, 2,\, 3$, leaving the other densities fixed. To this aim, we define $\rho_{t_1}^{12} := \rho_{t_1}^1 + \rho_{t_1}^2 = \chi_{J_1\cup J_2}\mathscr{L}^1$. In the time interval $[t_1, t_2]$, we leave fixed $\rho_0^k := \chi_{J_k}\mathscr{L}^1$ for $k \geq 4$ (if such k exist), and we apply again the above construction in $J_1 \cup J_2 \cup J_3$ to $\rho_{t_1}^{12}$ and $\rho_{t_1}^3 = \chi_{J_3}\mathscr{L}^1$. In this way, on $[t_1, t_2]$, $\rho_{t_1}^{12}$ is connected to $\rho_{t_2}^{12} := \frac{l_1+l_2}{l_1+l_2+l_3}\chi_{J_1\cup J_2\cup J_3}\mathscr{L}^1$, and $\rho_{t_1}^3$ is connected to $\rho_{t_2}^3 := \frac{l_3}{l_1+l_2+l_3}\chi_{J_1\cup J_2\cup J_3}\mathscr{L}^1$. Finally, it suffices to define $\rho_t^1 := \frac{l_1}{l_1+l_2}\rho_t^{12}$ and $\rho_t^2 := \frac{l_2}{l_1+l_2}\rho_t^{12}$.

In the third step we leave fixed the densities $\rho_{t_2}^k$ for $k \geq 5$, and we do the same construction as before adding the first three densities (that is, in this case one defines $\rho_{t_2}^{123} := \rho_{t_2}^1 + \rho_{t_2}^2 + \rho_{t_2}^3 = \chi_{J_1\cup J_2\cup J_3}\mathscr{L}^1$). In this way, we connect $\rho_{t_2}^{123}$ to $\rho_{t_3}^{123} := \frac{l_1+l_2+l_3}{l_1+l_2+l_3+l_4}\chi_{J_1\cup J_2\cup J_3\cup J_4}\mathscr{L}^1$ and $\rho_{t_2}^4$ to $\rho_{t_3}^4 := \frac{l_4}{l_1+l_2+l_3+l_4}\chi_{J_1\cup J_2\cup J_3\cup J_4}\mathscr{L}^1$, and then we define $\rho_t^k := \frac{l_k}{l_1+l_2+l_3}\rho_t^{123}$ for $k = 1,\, 2,\, 3$.

Iterating this construction on $[t_i, t_{i+1}]$ for $i \geq 4$, one obtains the desired maps $t \mapsto \rho_t^k$. Indeed, by construction $\rho_t^k \geq l_k$ on $J_{k,t}$, and $\sum_{k=1}^M \rho_t^k = 1$. Moreover, by (3.5.9), it is simple to see that in each time interval $[t_i, t_{i+1}]$ one has the bound

$$\operatorname{Lip}(\rho_\cdot^k) \leq \frac{M-1}{2}.$$

So the energy can be easily bounded by $1/\bar{b}^2\big(\frac{(M-1)^2}{16} + 1\big) \leq M^2/\bar{b}^2$. □

Proof. (of Theorem 3.5.2) By applying Theorem 3.5.3 to the optimal $\boldsymbol{\eta}$ connecting $\boldsymbol{i}$ to γ, we can find maps $g_k \in S(D)$ such that $\gamma_{g_k} \to \gamma$ narrowly and

$$\limsup_{k\to\infty} \overline{\delta}(\gamma_{\boldsymbol{i}}, \gamma_{g_k}) \leq \overline{\delta}(\gamma_{\boldsymbol{i}}, \gamma).$$

Now, if $d \geq 3$ we can use (3.3.12), the triangle inequality, and the density of SDiff(D) in $S(D)$ in the L^2 norm, to find maps $h_k \in \mathrm{SDiff}(D)$ such that

$$\limsup_{k\to\infty} \overline{\delta}(\gamma_{\boldsymbol{i}}, \gamma_{h_k}) \leq \overline{\delta}(\gamma_{\boldsymbol{i}}, \gamma)$$

and $\gamma_{h_k} \to \gamma$ narrowly. This gives the thesis. □

3.6 Necessary and sufficient optimality conditions

In this section we study necessary and sufficient optimality conditions for the generalized geodesics; we shall work mainly with the Lagrangian model, but we will use the equivalent Eulerian-Lagrangian model to transfer regularity informations for the pressure field to the Lagrangian model. Without any loss of generality, we assume throughout this section that $T = 1$.

The pressure field p can be identified, at least as a distribution (precisely, an element of the dual of $C^1([0,1] \times D)$), by the so-called dual least action principle introduced in [33]. In order to describe it, let us build a natural class of first variations in the Lagrangian model: given a smooth vector field $\boldsymbol{w}(t,x)$, vanishing for t sufficiently close to 0 and 1, we may define the maps $S^\varepsilon : \tilde{\Omega}(D) \to \tilde{\Omega}(D)$ by

$$S^\varepsilon(\omega, a)(t) := (e^{\varepsilon \boldsymbol{w}_t}\omega(t), a)\,, \tag{3.6.1}$$

where $e^{\varepsilon \boldsymbol{w}_t}x$ is the flow, in the (ε, x) variables, generated by the autonomous field $\boldsymbol{w}_t(x) = \boldsymbol{w}(t,x)$ (i.e. $e^{0\boldsymbol{w}_t} = \boldsymbol{i}$ and $\frac{d}{d\varepsilon}e^{\varepsilon \boldsymbol{w}_t}x = \boldsymbol{w}(t, e^{\varepsilon \boldsymbol{w}_t}x)$), and the perturbed generalized flows $\boldsymbol{\eta}_\varepsilon := (S^\varepsilon)_\#\boldsymbol{\eta}$. Notice that $\boldsymbol{\eta}_\varepsilon$ is incompressible if $\mathrm{div}\,\boldsymbol{w}_t = 0$, and more generally the density $\rho^{\boldsymbol{\eta}_\varepsilon}$ satisfies for all times $t \in (0,1)$ the continuity equation

$$\frac{d}{d\varepsilon}\rho^{\boldsymbol{\eta}_\varepsilon}(t,x) + \mathrm{div}(\boldsymbol{w}_t(x)\rho^{\boldsymbol{\eta}_\varepsilon}(t,x)) = 0. \tag{3.6.2}$$

This motivates the following definition.

Definition 3.6.1 (Almost incompressible flows). *We say that a probability measure $\boldsymbol{\nu}$ on $\Omega(D)$ is a almost incompressible generalized flow if $\rho^{\boldsymbol{\nu}} \in C^1([0,1] \times D)$ and*

$$\|\rho^{\boldsymbol{\nu}} - 1\|_{C^1([0,1]\times D)} \leq \frac{1}{2}.$$

Now we provide a slightly simpler proof of the characterization given in [33] of the pressure field (the original proof therein involved a time discretization argument).

Theorem 3.6.2. *For all η, $\gamma \in \Gamma(D)$ there exists $p \in [C^1([0,1]\times D)]^*$ such that*

$$\langle p, \rho^{\boldsymbol{\nu}} - 1\rangle_{(C^1)^*, C^1} \leq \mathscr{A}_1(\boldsymbol{\nu}) - \overline{\delta}^2(\eta,\gamma) \tag{3.6.3}$$

for all almost incompressible flows $\boldsymbol{\nu}$ satisfying (3.3.5).

Proof. Let us define the closed convex set $C := \{\rho \in C^1([0,1]\times D) : \|\rho - 1\|_{C^1} \leq \frac{1}{2}\}$, and the function $\phi : C^1([0,1]\times D) \to \mathbb{R}^+ \cup \{+\infty\}$ given by

$$\phi(\rho) := \begin{cases} \inf\{\mathscr{A}_1(\boldsymbol{\nu}) : \ \rho^{\boldsymbol{\nu}} = \rho \text{ and (3.3.5) holds}\} & \text{if } \rho \in C; \\ +\infty & \text{otherwise.} \end{cases}$$

We observe that $\phi(1) = \overline{\delta}^2(\eta,\gamma)$. Moreover, it is a simple exercise to prove that ϕ is convex and lower semicontinuous in $C^1([0,1]\times D)$. Let us now prove that ϕ has bounded (descending) slope at 1, i.e.

$$\limsup_{\rho\to 1} \frac{[\phi(1)-\phi(\rho)]^+}{\|1-\rho\|_{C^1}} < +\infty,$$

By [33, Proposition 2.1] we know that there exist $0 < \varepsilon < \frac{1}{2}$ and $c > 0$ such that, for any $\rho \in C$ with $\|\rho - 1\|_{C^1} \leq \varepsilon$, there is a Lipschitz family of diffeomorphisms $g_\rho(t,\cdot) : D \to D$ such that

$$g_\rho(t,\cdot)_\# \mu_D = \rho(t,\cdot)\mu_D,$$

$g_\rho(t,\cdot) = \boldsymbol{i}$ for $t = 0$, 1, and the Lipschitz constant of $(t,x) \mapsto g_\rho(t,x) - x$ is bounded by c. Thus, adapting the construction in [33, Proposition 2.1] (made for probability measures in $\Omega(D)$, and not in $\tilde{\Omega}(D)$), for any incompressible flow $\boldsymbol{\eta}$ connecting η to γ, and any $\rho \in C$, we can define an almost incompressible flow $\boldsymbol{\nu}$ still connecting η to γ such that $\rho^{\boldsymbol{\nu}} = \rho$, and

$$\mathscr{A}_1(\boldsymbol{\nu}) \leq \mathscr{A}_1(\boldsymbol{\eta}) + c'\|\rho - 1\|_{C^1}(1 + \mathscr{A}_1(\boldsymbol{\eta})),$$

where c' depends only on c (for instance, we define $\boldsymbol{\nu} := G_\# \boldsymbol{\eta}$, where $G : \tilde{\Omega}(D) \to \tilde{\Omega}(D)$ is the map induced by g_ρ via the formula $(\omega(t), a) \mapsto (g_\rho(t,\omega(t)), a)$). In particular, considering an optimal $\boldsymbol{\eta}$, we get

$$\phi(\rho) \leq \phi(1) + c\|\rho - 1\|_{C^1}(1 + \overline{\delta}^2(\eta,\gamma)) \tag{3.6.4}$$

for any $\rho \in C$ with $\|\rho - 1\|_{C^1} \leq \varepsilon$. This fact implies that ϕ is bounded on a neighbourhood of 1 in C. Now, it is a standard fact of convex analysis that a convex function bounded on a convex set is locally Lipschitz on that set. This provides the bounded slope property. By a simple application of the Hahn-Banach theorem (see for instance Proposition 1.4.4 in [11]), it follows that the subdifferential of ϕ at 1 is not empty, that is, there exists p in the dual of C^1 such that

$$\langle p, \rho - 1\rangle_{(C^1)^*, C^1} \leq \phi(\rho) - \phi(1).$$

This is indeed equivalent to (3.6.3). ∎

This result tells us that, if $\boldsymbol{\eta}$ is an optimal incompressible generalized flow connecting η to γ (i.e. $\mathscr{A}_1(\boldsymbol{\eta}) = \overline{\delta}^2(\eta,\gamma)$), and if we consider the augmented action

$$\mathscr{A}_1^p(\boldsymbol{\nu}) := \int_{\tilde\Omega(D)} \int_0^1 \frac{1}{2}|\dot\omega(t)|^2 \, dt \, d\boldsymbol{\nu}(\omega,a) - \langle p, \rho_{\boldsymbol{\nu}} - 1\rangle, \tag{3.6.5}$$

then $\boldsymbol{\eta}$ minimizes the new action among all almost incompressible flows $\boldsymbol{\nu}$ between η and γ.

Then, using the identities

$$\frac{d}{d\varepsilon}\frac{d}{dt} S^\varepsilon(\omega)(t)\bigg|_{\varepsilon=0} = \frac{d}{dt}\boldsymbol{w}(t,\omega(t)) = \partial_t \boldsymbol{w}(t,\omega(t)) + \nabla_x \boldsymbol{w}(t,\omega(t))\cdot\dot\omega(t)$$

and the convergence in the sense of distributions (ensured by (3.6.2)) of $(\rho^{\boldsymbol{\eta}_\varepsilon} - 1)/\varepsilon$ to $-\mathrm{div}\,\boldsymbol{w}$ as $\varepsilon \downarrow 0$, we obtain

$$0 = \frac{d}{d\varepsilon}\mathscr{A}_1^p(\boldsymbol{\eta}_\varepsilon)\bigg|_{\varepsilon=0} = \int_{\tilde\Omega(D)} \int_0^1 \dot\omega(t)\cdot \frac{d}{dt}\boldsymbol{w}(t,\omega(t)) \, dt \, d\boldsymbol{\eta}(\omega,a) + \langle p, \mathrm{div}\,\boldsymbol{w}\rangle. \tag{3.6.6}$$

As noticed in [33], this equation identifies uniquely the pressure field p (as a distribution) up to trivial modifications, i.e. additive perturbations depending on time only.

In the Eulerian-Lagrangian model, instead, the pressure field is defined (see (2.20) in [35]) and uniquely determined, still up to trivial modifications, by

$$\nabla p(t,x) = -\partial_t \left(\int_D \boldsymbol{v}(t,x,a)\, dc_{t,x}(a)\right) - \mathrm{div}\left(\int_D \boldsymbol{v}(t,x,a)\otimes\boldsymbol{v}(t,x,a)\, dc_{t,x}(a)\right), \tag{3.6.7}$$

all derivatives being understood in the sense of distributions in $(0,1)\times D$ (here $(c,\boldsymbol{v})$ is any optimal pair for the Eulerian-Lagrangian model). We used the same letter p to denote the pressure field in the two models: indeed, we have seen in the proof of Theorem 3.4.1 that, writing $\boldsymbol{\eta} = \boldsymbol{\eta}_a \otimes \mu_D$, the correspondence

$$\boldsymbol{\eta} \mapsto (c^{\boldsymbol{\eta}}_{t,a}, \boldsymbol{v}^{\boldsymbol{\eta}}_{t,a}) \qquad \text{with} \qquad c^{\boldsymbol{\eta}}_{t,a} := (e_t)_\#\boldsymbol{\eta}_a,\ \boldsymbol{v}^{\boldsymbol{\eta}}_{t,a} c^{\boldsymbol{\eta}}_{t,a} := (e_t)_\#(\dot\omega(t)\boldsymbol{\eta}_a)$$

maps optimal solutions for the first problem into optimal solutions for the second one. Since under this correspondence (3.6.7) reduces to (3.6.6), the two pressure fields coincide.

The following crucial regularity result for the pressure field is proved in the last section, and it improves in the time variable the regularity $\partial_{x_i} p \in \mathcal{M}_{\mathrm{loc}}\big((0,1)\times D\big)$ obtained by Brenier in [35].

Theorem 3.6.3 (Regularity of pressure). *Let* $(c, \boldsymbol{v})$ *be an optimal pair for the Eulerian-Lagrangian model, and let* p *be the pressure field identified by* (3.6.7). *Then* $\partial_{x_i} p \in L^2_{\mathrm{loc}}\left((0,1); \mathcal{M}(D)\right)$ *and*

$$p \in L^2_{\mathrm{loc}}\big((0,1); BV_{\mathrm{loc}}(D)\big) \subset L^2_{\mathrm{loc}}\big((0,1); L^{d/(d-1)}_{\mathrm{loc}}(D)\big).$$

In the case $D = \mathbb{T}^d$ *the same properties hold globally in space, i.e. replacing* $BV_{\mathrm{loc}}(D)$ *with* $BV(\mathbb{T}^d)$ *and* $L^{d/(d-1)}_{\mathrm{loc}}(D)$ *with* $L^{d/(d-1)}(\mathbb{T}^d)$.

The L^1_{loc} integrability of p allows much stronger variations in the Lagrangian model, that give rise to possibly nonsmooth densities, which may even vanish.

From now one we shall confine our discussion to the case of the flat torus $\mathbb{T}^d$, as our arguments involve some *global* smoothing that becomes more technical, and needs to be carefully checked in more general situations. We also set $\mu_{\mathbb{T}} = \mu_{\mathbb{T}^d}$ and denote by $d_{\mathbb{T}}$ the Riemannian distance in $\mathbb{T}^d$ (i.e. the distance modulo 1 in $\mathbb{R}^d/\mathbb{Z}^d$). In the next theorem we consider generalized flows $\boldsymbol{\nu}$ with *bounded compression*, defined by the property $\rho^{\boldsymbol{\nu}} \in L^\infty\left((0,1) \times D\right)$.

Theorem 3.6.4. *Let* $\boldsymbol{\eta}$ *be an optimal incompressible flow in* $\mathbb{T}^d$ *between* η *and* γ. *Then*

$$\langle p, \rho^{\boldsymbol{\nu}} - 1\rangle \le \mathscr{A}_1(\boldsymbol{\nu}) - \mathscr{A}_1(\boldsymbol{\eta}) \tag{3.6.8}$$

for any generalized flow with bounded compression $\boldsymbol{\nu}$ *between* η *and* γ *such that*

$$\rho^{\boldsymbol{\nu}}(t,\cdot) = 1 \text{ for } t \text{ sufficiently close to } 0,\ 1. \tag{3.6.9}$$

If $p \in L^1([0,1] \times \mathbb{T}^d)$, *the condition* (3.6.9) *is not required for the validity of* (3.6.8).

Proof. Let $J := \{\rho^{\boldsymbol{\nu}}(t,\cdot) \neq 1\} \Subset (0,1)$ and let us first assume that $\rho^{\boldsymbol{\nu}}$ is smooth. If $\|\rho^{\boldsymbol{\nu}} - 1\|_{C^1} \le 1/2$, then the result follows by Theorem 3.6.2. If not, for $\varepsilon > 0$ small enough $(1-\varepsilon)\boldsymbol{\eta} + \varepsilon\boldsymbol{\nu}$ is a slightly compressible generalized flow in the sense of Definition 3.6.1. Thus, we have

$$\varepsilon\langle p, \rho^{\boldsymbol{\nu}} - 1\rangle = \langle p, \rho^{(1-\varepsilon)\boldsymbol{\eta}+\varepsilon\boldsymbol{\nu}} - 1\rangle \le \mathscr{A}_1((1-\varepsilon)\boldsymbol{\eta} + \varepsilon\boldsymbol{\nu}) - \mathscr{A}_1(\boldsymbol{\eta}) = \varepsilon\left(\mathscr{A}_1(\boldsymbol{\nu}) - \mathscr{A}_1(\boldsymbol{\eta})\right),$$

and this proves the statement whenever $\rho^{\boldsymbol{\nu}}$ is smooth.

If $\rho^{\boldsymbol{\nu}}$ is not smooth, we need a regularization argument. Let us assume first that $\rho^{\boldsymbol{\nu}}$ is smooth in time, uniformly with respect to x, but not in space. We fix a cut-off function $\chi \in C^1_c(0,1)$ strictly positive on a neighbourhood of J and define, for $y \in \mathbb{R}^d$, the maps $T_{\varepsilon,y} : \tilde{\Omega}(\mathbb{T}^d) \to \tilde{\Omega}(\mathbb{T}^d)$ by

$$T_{\varepsilon,y}(\omega, a) := (\omega + \varepsilon y \chi, a), \qquad (\omega, a) \in \Omega(\mathbb{T}^d).$$

Then, we set $\boldsymbol{\nu}_\varepsilon := \int_{\mathbb{R}^d}(T_{\varepsilon,y})_\#\boldsymbol{\nu}\phi(y)\,dy$, where $\phi:\mathbb{R}^d\to[0,+\infty)$ is a standard convolution kernel. It is easy to check that $\boldsymbol{\nu}_\varepsilon$ still connects η to γ, and that

$$\rho^{\boldsymbol{\nu}_\varepsilon}(t,\cdot)=\rho^{\boldsymbol{\nu}}(t,\cdot)*\phi_{\varepsilon\chi(t)}\qquad\forall t\in[0,1],$$

where $\phi_\varepsilon(x)=\varepsilon^{-d}\phi(x/\varepsilon)$. Since

$$\lim_{\varepsilon\downarrow 0}\mathscr{A}_1(\boldsymbol{\nu}_\varepsilon)=\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^d}\int_{\tilde\Omega(\mathbb{T}^d)}\int_0^1|\dot\omega(t)+\varepsilon y\dot\chi(t)|^2\,dt\,d\boldsymbol{\nu}(\omega,a)\phi(y)\,dy=\mathscr{A}_1(\boldsymbol{\nu})$$

we can pass to the limit in (3.6.8) with $\boldsymbol{\nu}_\varepsilon$ in place of $\boldsymbol{\nu}$, which are smooth.

In the general case we fix a convolution kernel with compact support $\varphi(t)$ and, with the same choice of χ done before, we define the maps

$$T_\varepsilon(\omega,a)(t):=(\int_0^1\omega(t-s\varepsilon\chi(t))\varphi(s)\,ds,a).$$

Setting $\boldsymbol{\nu}_\varepsilon=(T_\varepsilon)_\#\boldsymbol{\nu}$, it is easy to check that $\mathscr{A}_1(\boldsymbol{\nu}_\varepsilon)\to\mathscr{A}_1(\boldsymbol{\nu})$ and that

$$\rho^{\boldsymbol{\nu}_\varepsilon}(t,x)=\int_0^1\rho^{\boldsymbol{\nu}}(t-s\varepsilon\chi(t),x)\varphi(s)\,ds$$

are smooth in time, uniformly in x. So, by applying (3.6.8) with $\boldsymbol{\nu}_\varepsilon$ in place of $\boldsymbol{\nu}$, we obtain the inequality in the limit.

Finally, if p is globally integrable, we can approximate any generalized flow with bounded compression $\boldsymbol{\nu}$ between η and γ by transforming ω into $\omega\circ\psi_\varepsilon$, where $\psi_\varepsilon:[0,1]\to[0,1]$ is defined by $\psi_\varepsilon(t):=\frac{1}{1-2\varepsilon}\int_0^t\chi_{[\varepsilon,1-\varepsilon]}(s)ds$ (so that ψ_ε is constant for t close to 0 and 1). Passing to the limit as $\varepsilon\downarrow 0$ we obtain the inequality even without the condition $\rho^{\boldsymbol{\nu}}(t,\cdot)=1$ for t close to 0, 1. □

Remark 3.6.5 (Smoothing of flows and plans). Notice that the same smoothing argument can be used to prove this statement: given a flow $\boldsymbol{\eta}$ between $\eta=\eta_a\otimes\mu_{\mathbb{T}}$ and $\gamma=\gamma_a\otimes\mu_{\mathbb{T}}$ (not necessarily with bounded compression), we can find flows with bounded compression $\boldsymbol{\eta}^\varepsilon$ connecting $\eta^\varepsilon:=(\eta_a)*\phi_\varepsilon\otimes\mu_{\mathbb{T}}$ to $\gamma^\varepsilon:=(\gamma_a)*\phi_\varepsilon\otimes\mu_{\mathbb{T}}$, with $\mathscr{A}_T(\boldsymbol{\eta}^\varepsilon)=\mathscr{A}_T(\boldsymbol{\eta})$ and

$$\int_{\tilde\Omega(\mathbb{T}^d)}\int_0^1 r_\varepsilon(\tau,\omega)\,d\tau\,d\boldsymbol{\eta}(\omega,a)=\int_{\tilde\Omega(\mathbb{T}^d)}\int_0^1 r(\tau,\omega)\,d\tau\,d\boldsymbol{\eta}^\varepsilon(\omega,a)\qquad\forall r\in L^1\left([0,1]\times\mathbb{T}^d\right)$$

(where, as usual, $r_\varepsilon(t,x)=r(t,\cdot)*\phi_\varepsilon(x)$). In order to have these properties, it suffices to define

$$\boldsymbol{\eta}^\varepsilon:=\int_{\mathbb{R}^d}(\sigma_{\varepsilon y})_\#\boldsymbol{\eta}\,\phi(y)\,dy,$$

where $\sigma_z(\omega,a)=(\omega+z,a)$. Notice also that the "mollified plans" η^ε, γ^ε converge to η, γ in $(\Gamma(\mathbb{T}^d),\overline{\delta})$: if we consider the map $S^\varepsilon_y:\mathbb{T}^d\to\Omega(\mathbb{T}^d)$ given by $x\mapsto\omega_x(t):=x+\varepsilon ty$, the generalized incompressible flow $\boldsymbol{\nu}^\varepsilon=\boldsymbol{\nu}^\varepsilon_a\otimes\mu_{\mathbb{T}}$, with

$$\boldsymbol{\nu}^\varepsilon_a:=\int_{\mathbb{R}^d}(S^\varepsilon_y)_\#\lambda_a\,\phi(y)\,dy,$$

connects in $[0,1]$ the plan $\lambda=\lambda_a\otimes\mu_{\mathbb{T}}$ to $\lambda^\varepsilon=(\lambda_a*\phi_\varepsilon)\otimes\mu_{\mathbb{T}}$, with an action equal to $\varepsilon^2\int_{\mathbb{R}^d}|y|^2\phi(y)\,dy$.

In order to state necessary and sufficient optimality conditions at the level of single fluid paths, we have to take into account that the pressure field is not pointwise defined, and to choose a particular representative in its equivalence class, modulo negligible sets in spacetime. Henceforth, we define

$$\bar p(t,x):=\liminf_{\varepsilon\downarrow 0}p_\varepsilon(t,x), \tag{3.6.10}$$

where, thinking of $p(t,\cdot)$ as a 1-periodic function in $\mathbb{R}^d$, p_ε is defined by

$$p_\varepsilon(t,x):=(2\pi)^{-d/2}\int_{\mathbb{R}^d}p(t,x+\varepsilon y)e^{-|y|^2/2}\,dy.$$

Notice that p_ε is smooth and still 1-periodic. The choice of the heat kernel here is convenient, because of the semigroup property $p_{\varepsilon+\varepsilon'}=(p_\varepsilon)_{\varepsilon'}$. Recall that $\bar p$ is a representative, because at any Lebesgue point x of $p(t,\cdot)$ the limit of $p_\varepsilon(t,x)$ exists, and coincides with $p(t,x)$.

In order to handle passages to limits, we need also uniform pointwise bounds on p_ε; therefore we define

$$Mf(x):=\sup_{\varepsilon>0}\,(2\pi)^{-d/2}\int_{\mathbb{R}^d}|f|(x+\varepsilon y)e^{-|y|^2/2}\,dy,\qquad f\in L^1(\mathbb{T}^d). \tag{3.6.11}$$

We will use the following facts: first,

$$Mf_\varepsilon=\sup_{\varepsilon'>0}|f_\varepsilon|_{\varepsilon'}\le\sup_{\varepsilon'>0}(|f|_\varepsilon)_{\varepsilon'}\le\sup_{r>0}|f|_r=Mf$$

because of the semigroup property; second, standard maximal inequalities imply $\|Mf\|_{L^p(\mathbb{T}^d)}\le c_p\|f\|_{L^p(\mathbb{T}^d)}$ for all $p>1$. Setting $Mp(t,x):=Mp(t,\cdot)(x)$, by Theorem 3.6.3 we infer that $Mp\in L^2_{\rm loc}\big((0,1),L^{d/d-1}(\mathbb{T}^d)\big)$, so that in particular $Mp\in L^1_{\rm loc}\big((0,1)\times\mathbb{T}^d\big)$. This is the integrability assumption on p that will play a role in the rest of this section.

Definition 3.6.6 (q-minimizing path). *Let $\omega \in H^1((0,1);D)$ with $Mq(\tau,\omega) \in L^1(0,1)$. We say that ω is a* q-minimizing path *if*

$$\int_0^1 \frac{1}{2}|\dot\omega(\tau)|^2 - q(\tau,\omega)\,d\tau \le \int_0^1 \frac{1}{2}|\dot\omega(\tau)+\dot\delta(\tau)|^2 - q(\tau,\omega+\delta)\,d\tau$$

for all $\delta \in H^1_0((0,1);D)$ with $Mq(\tau,\omega+\delta) \in L^1(0,1)$.
Analogously, we say that ω is a locally q-minimizing path *if*

$$\int_s^t \frac{1}{2}|\dot\omega(\tau)|^2 - q(\tau,\omega)\,d\tau \le \int_s^t \frac{1}{2}|\dot\omega(\tau)+\dot\delta(\tau)|^2 - q(\tau,\omega+\delta)\,d\tau \tag{3.6.12}$$

for all $[s,t] \subset (0,1)$ and all $\delta \in H^1_0((s,t);D)$ with $Mq(\tau,\omega+\delta) \in L^1(s,t)$.

Remark 3.6.7. We notice that, for incompressible flows $\boldsymbol{\eta}$, the L^1 (resp. $L^1_{\rm loc}$) integrability of $Mq(\tau,\omega)$ imposed on the curves ω (and on their perturbations $\omega+\delta$) is satisfied $\boldsymbol{\eta}$-a.e. if $Mq \in L^1\big((0,1)\times\mathbb{T}^d\big)$ (resp. $Mq \in L^1_{\rm loc}\big((0,1)\times\mathbb{T}^d\big)$); this can simply be obtained first noticing that the incompressibility of $\boldsymbol{\eta}$ and Fubini's theorem give

$$\int_{\tilde\Omega(\mathbb{T}^d)}\int_J f(\tau,\omega)\,d\tau\,d\boldsymbol{\eta}(\omega,a) = \int_J\int_{\mathbb{T}^d} f(\tau,x)\,d\mu_{\mathbb{T}^d}(x)\,d\tau$$

for all nonnegative Borel functions f and all intervals $J \subset (0,1)$, and then applying this identity to $f = Mq$.

Theorem 3.6.8 (First necessary condition). *Let $\boldsymbol{\eta} = \boldsymbol{\eta}_a \otimes \mu_{\mathbb{T}}$ be any optimal incompressible flow on $\mathbb{T}^d$. Then, $\boldsymbol{\eta}$ is concentrated on locally $\bar p$-minimizing paths, where $\bar p$ is the precise representative of the pressure field p, and on $\bar p$-minimizing paths if $Mp \in L^1([0,1]\times\mathbb{T}^d)$.*

Proof. With no loss of generality we identify $\mathbb{T}^d$ with $\mathbb{R}^d/\mathbb{Z}^d$. Let $\boldsymbol{\eta}$ be an optimal incompressible flow and $[s,t] \subset (0,1)$. We fix a nonnegative function $\chi \in C^1_c(0,1)$ with $\{\chi > 0\} = (s,t)$. Given $\delta \in H^1_0\big([s,t];\mathbb{T}^d\big)$, $y \in \mathbb{R}^d$ and a Borel set $E \subset \tilde\Omega(\mathbb{T}^d)$, we define $T_{\varepsilon,y} : \tilde\Omega(\mathbb{T}^d) \to \tilde\Omega(\mathbb{T}^d)$ by

$$T_{\varepsilon,y}(\omega,a) := \begin{cases} (\omega,a) & \text{if } \omega \notin E;\\ (\omega+\delta+\varepsilon y\chi, a) & \text{if } \omega \in E\end{cases}$$

(of course, the sum is understood modulo 1) and $\boldsymbol{\nu}_{\varepsilon,y} := (T_{\varepsilon,y})_\#\boldsymbol{\eta}$.

It is easy to see that $\boldsymbol{\nu}_{\varepsilon,y}$ is a flow with bounded compression, since for all times τ the curves $\omega(\tau)$ are either left unchanged, or translated by the constant $\delta(\tau)+\varepsilon y\chi(\tau)$, so that the density produced by $\boldsymbol{\nu}_{\varepsilon,y}$ is at most 2, and equal to 1 outside the interval $[s,t]$.

Therefore, by Theorem 3.6.4 we get

$$\int_{\mathbb{T}^d}\int_s^t \bar{p}(\rho^{\boldsymbol{\nu}_{\varepsilon,y}}-1)\,d\tau\,d\mu_{\mathbb{T}} \le \int_E \mathscr{A}_1(\omega+\delta+\varepsilon y\chi)-\mathscr{A}_1(\omega)\,d\boldsymbol{\eta}(\omega,a).$$

Rearranging terms, we get

$$\int_E\int_s^t \frac{1}{2}|\dot\omega|^2-\bar{p}(\tau,\omega)\,d\tau\,d\boldsymbol{\eta}(\omega,a) \le \int_E\left[\int_s^t \frac{1}{2}|\dot\omega+\dot\delta+\varepsilon y\dot\chi|^2-\bar{p}(\tau,\omega+\delta+\varepsilon y\chi)\,d\tau\right]d\boldsymbol{\eta}(\omega,a).$$

We can now average the above inequality using the heat kernel $\phi(y)=(2\pi)^{-d/2}e^{-|y|^2/2}$, and we obtain

$$\begin{aligned}\int_E\int_s^t \frac{1}{2}|\dot\omega|^2-\bar{p}(\tau,\omega)\,d\tau\,d\boldsymbol{\eta}(\omega,a)&\\ \le \int_E\left[\int_{\mathbb{R}^d}\int_s^t \frac{1}{2}|\dot\omega+\dot\delta+\varepsilon y\dot\chi|^2\,d\tau\phi(y)\,dy-\int_s^t p_{\varepsilon\chi(\tau)}(\tau,\omega+\delta)\,d\tau\right]d\boldsymbol{\eta}(\omega,a).&\end{aligned}$$

Now, let $\mathcal{D}\subset H_0^1\left([s,t];\mathbb{T}^d\right)$ be a countable dense subset; by the arbitrariness of E and Remark 3.6.7 we infer the existence of a $\boldsymbol{\eta}$-negligible Borel set $B\subset\tilde\Omega(\mathbb{T}^d)$ such that $Mp(\tau,\omega)\in L^1(s,t)$ and

$$\int_s^t \frac{1}{2}|\dot\omega|^2-\bar{p}(\tau,\omega)\,d\tau \le \int_{\mathbb{R}^d}\int_s^t \frac{1}{2}|\dot\omega+\dot\delta+\varepsilon y\dot\chi|^2\,d\tau\phi(y)\,dy-\int_s^t p_{\varepsilon\chi(\tau)}(\tau,\omega+\delta)\,d\tau$$

holds for all $\varepsilon=1/n$, $\delta\in\mathcal{D}$ and $(\omega,a)\in\tilde\Omega(\mathbb{T}^d)\setminus B$. By a density argument, we see that the same inequality holds for all $\varepsilon=1/n$, $\delta\in H_0^1\left([s,t];\mathbb{T}^d\right)$, and $\omega\in\tilde\Omega(\mathbb{T}^d)\setminus B$.

Now, if $Mp(\tau,\omega+\delta)\in L^1(s,t)$, since $\delta\in H_0^1\left([s,t];\mathbb{T}^d\right)$ we have that $Mp(\tau,\omega+\delta)\in L^1(s,t)$, and we can use the bound $|p_\varepsilon|\le Mp$ to pass to the limit as $\varepsilon\downarrow 0$ to obtain that (3.6.12) holds with $q=\bar{p}$.

The proof of the global minimality property in the case when $p\in L^1([0,T]\times\mathbb{T}^d)$ is similar, just letting δ vary in $H_0^1\left([0,1];\mathbb{T}^d\right)$ and using a fixed function $\chi\in C^1([0,1])$ with $\chi(0)=\chi(1)=0$ and $\chi>0$ in $(0,1)$. □

In order to state the second necessary optimality condition fulfilled by minimizers, we need some preliminary definition. Let $q\in L^1\left([s,t]\times D\right)$ and let us define the cost $c_q^{s,t}:D\times D\to\overline{\mathbb{R}}$ of the minimal connection in $[s,t]$ between x and y, namely

$$c_q^{s,t}(x,y):=\inf\left\{\int_s^t \frac{1}{2}|\dot\omega(\tau)|^2-q(\tau,\omega)\,d\tau:\ \omega(s)=x,\,\omega(t)=y,\,Mq(\tau,\omega)\in L^1(s,t)\right\},\tag{3.6.13}$$

with the convention $c_q^{s,t}(x,y) = +\infty$ if no admissible curve ω exists. Using this cost function $c_q^{s,t}$, we can consider the induced optimal transport problem, namely

$$W_{c_q^{s,t}}(\mu_1,\mu_2) := \inf\left\{\int_{D\times D} c_q^{s,t}(x,y)\,d\lambda(x,y):\ \lambda\in\Gamma(\mu_1,\mu_2),\ (c_q^{s,t})^+\in L^1(\lambda)\right\}, \tag{3.6.14}$$

where $\Gamma(\mu_1,\mu_2)$ is the family of all probability measures λ in $D\times D$ whose first and second marginals are respectively μ_1 and μ_2. Again, we set by convention $W_{c_q^{s,t}}(\mu_1,\mu_2)=+\infty$ if no admissible λ exists.

Unlike most classical situations (see [132]), existence of an optimal λ is not guaranteed because $c_q^{s,t}$ are not lower semicontinuous in $D\times D$, and also it seems difficult to get lower bounds on $c_q^{s,t}$. It will be useful, however, the following *upper* bound on $W_{c_q^{s,t}}$:

Lemma 3.6.9. *If $Mq\in L^1([s,t]\times\mathbb{T}^d)$ there exists a nonnegative $\mu_{\mathbb{T}}$-integrable function $K_q^{s,t}$ satisfying*

$$c_q^{s,t}(x,y)\le K_q^{s,t}(x)+K_q^{s,t}(y)\qquad \forall x,y\in\mathbb{T}^d. \tag{3.6.15}$$

Remark 3.6.10. By (3.6.15) we deduce that, if $K_q^{s,t}\in L^1(\mu_1+\mu_2)$, then $(c_q^{s,t})^+\in L^1(\lambda)$ for all $\lambda\in\Gamma(\mu_1,\mu_2)$ and we have

$$\int_{\mathbb{T}^d\times\mathbb{T}^d} c_q^{s,t}(x,y)\,d\lambda(x,y)\le\int_{\mathbb{T}^d}K_q^{s,t}(w)\,d(\mu_1+\mu_2)(w)\qquad\forall\lambda\in\Gamma(\mu_1,\mu_2).$$

In particular, $W_{c_q^{s,t}}(\mu_1,\mu_2)$ as defined in (3.6.14) is not equal to $+\infty$.

Proof. Assume $s=0$ and let $l=t/2$. Let us fix x, $y\in\mathbb{T}^d$; given $z\in\mathbb{T}^d$ we consider the projection on $\mathbb{T}^d$ of the Euclidean path

$$\omega_z(\tau):=\begin{cases} x+\frac{\tau}{l}(z-x) & \text{if } \tau\in[0,l];\\ z+\frac{\tau-l}{l}(y-z) & \text{if } \tau\in[l,t].\end{cases}$$

This path leads to the estimate

$$c_q^{0,t}(x,y)\le\frac{d_{\mathbb{T}}^2(x,z)+d_{\mathbb{T}}^2(z,y)}{2l}+\int_0^l Mq(\tau,x+\frac{\tau}{l}(z-x))\,d\tau+\int_l^t Mq(\tau,z+\frac{\tau-l}{l}(y-z))\,d\tau.$$

By integrating the free variable z with respect to $\mu_{\mathbb{T}}$, since $d_{\mathbb{T}}\le\frac{\sqrt{d}}{2}$ on $\mathbb{T}^d\times\mathbb{T}^d$, we get

$$c_q^{0,t}(x,y)\le\frac{d}{4l}+\int_{\mathbb{T}^d}\int_0^l Mq(\tau,x+\frac{\tau}{l}(z-x))+Mq(l+\tau,z+\frac{\tau}{l}(y-z))\,d\tau\,d\mu_{\mathbb{T}}(z).$$

Therefore, the function

$$K_q^{0,t}(w) := \frac{d}{4l} + \int_{\mathbb{T}^d} \int_0^l Mq(\tau, w + \frac{\tau}{l}(z-w)) + Mq(l+\tau, z+\frac{\tau}{l}(w-z))\, d\tau\, d\mu_{\mathbb{T}}(z) \quad (3.6.16)$$

fulfils (3.6.15). It is easy to check, using Fubini's theorem, that $K_q^{0,t}$ is $\mu_{\mathbb{T}}$-integrable in $\mathbb{T}^d$. Indeed,

$$\begin{aligned}
\int_{\mathbb{T}^d} K_q^{0,t}(w)\, d\mu_{\mathbb{T}}(w) &= \frac{d}{4l} + \int_{\mathbb{T}^d}\int_{\mathbb{T}^d}\int_0^l Mq(\tau, w + \frac{\tau}{l}(z-w))\, d\tau\, d\mu_{\mathbb{T}}(z)\, d\mu_{\mathbb{T}}(w) \\
&+ \int_{\mathbb{T}^d}\int_{\mathbb{T}^d}\int_0^l Mq(l+\tau, z + \frac{\tau}{l}(w-z))\, d\tau\, d\mu_{\mathbb{T}}(w)\, d\mu_{\mathbb{T}}(z) \\
&= \frac{d}{4l} + \int_{\mathbb{T}^d}\int_{\mathbb{T}^d}\int_0^l Mq(\tau, w + \frac{\tau}{l}y)\, d\tau\, d\mu_{\mathbb{T}}(y)\, d\mu_{\mathbb{T}}(w) \\
&+ \int_{\mathbb{T}^d}\int_{\mathbb{T}^d}\int_0^l Mq(l+\tau, z + \frac{\tau}{l}y)\, d\tau\, d\mu_{\mathbb{T}}(z)\, d\mu_{\mathbb{T}}(y) \\
&= \frac{d}{4l} + \int_0^l\int_{\mathbb{T}^d}\int_{\mathbb{T}^d} Mq(\tau, w + \frac{\tau}{l}y) + Mq(l+\tau, w + \frac{\tau}{l}y)\, d\mu_{\mathbb{T}}(w)\, d\mu_{\mathbb{T}}(y)\, d\tau \\
&= \frac{d}{4l} + \int_0^t\int_{\mathbb{T}^d} Mq(\tau, w)\, d\mu_{\mathbb{T}}(w)\, d\tau < +\infty.
\end{aligned}$$

$\square$

In the proof of the next theorem we are going to use the *measurable selection theorem* (see [43, Theorems III.22 and III.23]): if $(A, \mathcal{A}, \nu)$ is a measure space, X is a Polish space and $E \subset A \times X$ is $\mathcal{A}_\nu \otimes \mathcal{B}(X)$-measurable, where $\mathcal{A}_\nu$ is the ν-completion of $\mathcal{A}$, then:

(i) the projection $\pi_A(E)$ of E on A is $\mathcal{A}_\nu$-measurable;

(ii) there exists a $(\mathcal{A}_\nu, \mathcal{B}(X))$-measurable map $\sigma : \pi(E) \to X$ such that $(x, \sigma(x)) \in E$ for ν-a.e. $x \in \pi_A(E)$.

The next theorem will provide a new necessary optimality condition involving not only the path that should be followed between x and y (which, as we proved, should minimize the Lagrangian $\mathcal{L}_{\bar{p}}$ in (3.1.8)), but also the "weights" given to the paths. We observe that, when a variation of these weights is performed, new flows $\tilde{\boldsymbol{\eta}}$ between η and γ are built which need not be of bounded compression, for which $(e_t)_\#\tilde{\boldsymbol{\eta}}$ might be even singular with respect to $\mu_{\mathbb{T}}$; therefore we can't use directly them in the variational principle (3.6.8); however, this difficulty can be overcome by the smoothing procedure in Remark 3.6.5.

Theorem 3.6.11 (Second necessary condition). *Let $\boldsymbol{\eta} = \boldsymbol{\eta}_a \otimes \mu_{\mathbb{T}}$ be an optimal incompressible flow on $\mathbb{T}^d$ between η and γ. Then, for all intervals $[s,t] \subset (0,1)$, $W_{c_{\bar{p}}^{s,t}}(\eta_a, \gamma_a) \in \mathbb{R}$ and the plan $(e_s, e_t)_\# \boldsymbol{\eta}_a$ is optimal, relative to the cost $c_{\bar{p}}^{s,t}$ defined in* (3.6.13)*, for $\mu_{\mathbb{T}}$-a.e. a.*

Proof. Let $[s,t] \subset (0,1)$ be fixed. Since

$$\begin{aligned}\int_{\mathbb{T}^d}\int_{\mathbb{T}^d\times\mathbb{T}^d} c_{\bar{p}}^{s,t}(x,y)\, d(e_s,e_t)_\#\boldsymbol{\eta}_a\, d\mu_{\mathbb{T}}(a) \;&\le\; \int_{\mathbb{T}^d}\int_{\Omega(\mathbb{T}^d)}\int_s^t \frac{1}{2}|\dot\omega(\tau)|^2 - \bar{p}(\tau,\omega)\, d\tau d\boldsymbol{\eta}_a(\omega)\, d\mu_{\mathbb{T}}(a)\\ &=\; (t-s)\overline{\delta}^2(\eta,\gamma),\end{aligned} \tag{3.6.17}$$

it suffices to show that

$$(t-s)\overline{\delta}^2(\eta,\gamma) \le \int_{\mathbb{T}^d} W_{c_{\bar{p}}^{s,t}}(\eta_a^s, \gamma_a^t)\, d\mu_{\mathbb{T}}(a). \tag{3.6.18}$$

We are going to prove this fact by a smoothing argument. We set $\eta^s = \eta_a^s \otimes \mu_{\mathbb{T}}$, $\gamma^t = \gamma_a^t \otimes \mu_{\mathbb{T}}$, with $\eta_a^s = (e_s)_\#\boldsymbol{\eta}_a$, $\gamma_a^t = (e_t)_\#\boldsymbol{\eta}_a$. Recall that Remark 3.3.1 gives

$$\overline{\delta}(\eta,\eta^s) = s\overline{\delta}(\eta,\gamma), \qquad \overline{\delta}(\gamma^t,\gamma) = (1-t)\overline{\delta}(\eta,\gamma).$$

First, we notice that Lemma 3.6.9 gives

$$\begin{aligned}\int_{\mathbb{T}^d} W_{c_{-|\bar{p}|}^{s,t}}(\eta_a^s,\gamma_a^t)\, d\mu_{\mathbb{T}}(a) &\le \int_{\mathbb{T}^d}\int_{\mathbb{T}^d} K_{-|\bar{p}|}^{s,t}(w)\, d(\eta_a^s+\gamma_a^t)(w)\, d\mu_{\mathbb{T}}(a)\\ &= 2\int_{\mathbb{T}^d} K_{-|\bar{p}|}^{s,t}(w)\, d\mu_{\mathbb{T}}(w) < +\infty.\end{aligned} \tag{3.6.19}$$

We also remark that, since $\tau \mapsto \|p_\varepsilon(\tau,\cdot)\|_\infty$ is integrable in (s,t), for any $\varepsilon > 0$ the cost $c_{p_\varepsilon}^{s,t}$ is bounded both from above and below. Next, we show that

$$c_{\bar{p}}^{s,t}(x,y) \ge \limsup_{\varepsilon\downarrow 0} c_{p_\varepsilon}^{s,t}(x,y) \qquad \forall (x,y)\in\mathbb{T}^d\times\mathbb{T}^d. \tag{3.6.20}$$

Indeed, let $\omega \in H^1\left([s,t];\mathbb{T}^d\right)$ with $\omega(s) = x$, $\omega(t) = y$ and $Mp(\tau,\omega) \in L^1(s,t)$ (if there is no such ω, there is nothing to prove). By the pointwise bound $|p_\varepsilon| \le Mp$ and Lebesgue's theorem, we get

$$\int_s^t \frac{1}{2}|\dot\omega(\tau)|^2 - \bar{p}(\tau,\omega)\, d\tau = \lim_{\varepsilon\downarrow 0}\int_s^t \frac{1}{2}|\dot\omega(\tau)|^2 - p_\varepsilon(\tau,\omega)\, d\tau.$$

By the $L^1(L^\infty)$ bound on Mp_ε, the curve ω is admissible also for the variational problem defining $c_{p_\varepsilon}^{s,t}$, therefore the above limit provides an upper bound on $\limsup_\varepsilon c_{p_\varepsilon}^{s,t}(x,y)$. By minimizing with respect to ω we obtain (3.6.20).

By (3.6.19) and the pointwise bound $\bar p \geq -|\bar p|$ we infer that the positive part of $W_{c^{s,t}_{\bar p}}(\eta^s_a,\gamma^t_a)$ is $\mu_{\mathbb{T}}$-integrable. Let now $\delta>0$ be fixed, and let us consider the compact space $X := \mathscr{P}(\mathbb{T}^d\times\mathbb{T}^d)$ and the $\mathcal{B}(\mathbb{T}^d)_{\mu_{\mathbb{T}}}\otimes\mathcal{B}(X)$-measurable set

$$E := \left\{(a,\lambda)\in\mathbb{T}^d\times X:\ \lambda\in\Gamma(\eta^s_a,\gamma^t_a),\ \int_{\mathbb{T}^d\times\mathbb{T}^d} c^{s,t}_{\bar p}(x,y)\,d\lambda < \delta + \Big(W_{c^{s,t}_{\bar p}}(\eta^s_a,\gamma^t_a)\vee -\frac{1}{\delta}\Big)\right\}$$

(we skip the proof of the measurability, that is based on tedious but routine arguments). Since $W_{c^{s,t}_{\bar p}}(\eta^s_a,\gamma^t_a) < +\infty$ for $\mu_{\mathbb{T}}$-a.e. a, we obtain that for $\mu_{\mathbb{T}}$-a.e. $a\in\mathbb{T}^d$ there exists $\lambda\in\Gamma(\eta^s_a,\gamma^t_a)$ with $(a,\lambda)\in E$. Thanks to the measurable selection theorem we can select a Borel family $a\mapsto\lambda_a\in\mathscr{P}(\mathbb{T}^d\times\mathbb{T}^d)$ such that $\lambda_a\in\Gamma(\eta^s_a,\gamma^t_a)$ and

$$\int_{\mathbb{T}^d\times\mathbb{T}^d} c^{s,t}_{\bar p}(x,y)\,d\lambda_a < \delta + \Big(W_{c^{s,t}_{\bar p}}(\eta^s_a,\gamma^t_a)\vee -\frac{1}{\delta}\Big) \qquad \text{for } \mu_{\mathbb{T}}\text{-a.e. } a\in\mathbb{T}^d.$$

By Lemma 3.6.9 and Remark 3.6.10 we get

$$c^{s,t}_{p_\varepsilon}(x,y)\leq K^{s,t}_{p_\varepsilon}(x)+K^{s,t}_{p_\varepsilon}(y)\leq K^{s,t}_{p}(x)+K^{s,t}_{p}(y) \qquad \forall x,y\in\mathbb{T}^d$$

and

$$\int_{\mathbb{T}^d}\int_{\mathbb{T}^d\times\mathbb{T}^d} K^{s,t}_p(x)+K^{s,t}_p(y)\,d\lambda_a\,d\mu_{\mathbb{T}}(a) = \int_{\mathbb{T}^d}\int_{\mathbb{T}^d} K^{s,t}_p\,d(\eta^s_a+\gamma^t_a)\,d\mu_{\mathbb{T}}(a) < +\infty$$

(we used the pointwise bound $Mp_\varepsilon\leq Mp$ and the fact that $q\mapsto K^{s,t}_q$ has a monotone dependence upon Mq, see (3.6.16)). Therefore (3.6.20) and Fatou's lemma give

$$\delta + \int_{\mathbb{T}^d}\Big(W_{c^{s,t}_{\bar p}}(\eta^s_a,\gamma^t_a)\vee -\frac{1}{\delta}\Big)\,d\mu_{\mathbb{T}}(a) \geq \limsup_{\varepsilon\downarrow 0}\int_{\mathbb{T}^d}\int_{\mathbb{T}^d\times\mathbb{T}^d} c^{s,t}_{p_\varepsilon}(x,y)d\lambda_a\,d\mu_{\mathbb{T}}(a). \tag{3.6.21}$$

Still thanks to the measurable selection theorem, we can find a Borel map $(x,y,a)\mapsto \omega^{x,y}_{a,\varepsilon}\in C\left([s,t];\mathbb{T}^d\right)$ with $\omega^{x,y}_{a,\varepsilon}(s)=x$, $\omega^{x,y}_{a,\varepsilon}(t)=y$, $Mp_\varepsilon(\tau,\omega^{x,y}_{a,\varepsilon})\in L^1(s,t)$ and

$$\int_s^t \frac{1}{2}|\dot\omega^{x,y}_{a,\varepsilon}|^2 - p_\varepsilon(\tau,\omega^{x,y}_{a,\varepsilon})\,d\tau < \delta + c^{s,t}_{p_\varepsilon}(x,y) \qquad \text{for } \lambda_a\otimes\mu_{\mathbb{T}}\text{-a.e. } (x,y,a).$$

Let $\boldsymbol{\lambda}^\varepsilon = \boldsymbol{\lambda}^\varepsilon_a\otimes\mu_{\mathbb{T}}$ be the push-forward, under the map $(x,y,a)\mapsto\omega^{x,y}_{a,\varepsilon}$, of the measure $\lambda_a\otimes\mu_{\mathbb{T}}$; by construction this measure fulfils $(e_s)_\#\boldsymbol{\lambda}^\varepsilon_a=\eta^s_a$, $(e_t)_\#\boldsymbol{\lambda}^\varepsilon_a=\gamma^t_a$, (because the marginals of λ_a are η^s_a and γ^t_a), therefore it connects η^s to γ^t in $[s,t]$. Then, from (3.6.21) we get

$$2\delta+\int_{\mathbb{T}^d}\Big(W_{c^{s,t}_{\bar p}}(\eta^s_a,\gamma^t_a)\vee-\frac{1}{\delta}\Big)\,d\mu_{\mathbb{T}}(a)\geq\limsup_{\varepsilon\downarrow 0}\int_{C([s,t];\mathbb{T}^d)\times\mathbb{T}^d}\int_s^t\frac{1}{2}|\dot\omega(\tau)|^2-p_\varepsilon(\tau,\omega)\,d\tau\,d\boldsymbol{\lambda}^\varepsilon(\omega,a).$$

Eventually, Remark 3.6.5 provides us with a flow with bounded compression $\hat{\boldsymbol{\lambda}}^{\varepsilon}$ connecting $\eta^{s,\varepsilon}$ to $\gamma^{t,\varepsilon}$ in $[s,t]$ with

$$2\delta+\int_{\mathbb{T}^d}\Big(W_{c_{\bar{p}}^{s,t}}(\eta_a,\gamma_a)\vee-\frac{1}{\delta}\Big)\,d\mu_{\mathbb{T}}(a)\geq\limsup_{\varepsilon\downarrow 0}\int_{C([s,t];\mathbb{T}^d)\times\mathbb{T}^d}\int_s^t\frac{1}{2}|\dot\omega(\tau)|^2-\bar p(\tau,\omega)\,d\tau\,d\hat{\boldsymbol{\lambda}}^{\varepsilon}(\omega,a). \tag{3.6.22}$$

Since $\eta^{s,\varepsilon}\to\eta^s$ and $\gamma^{t,\varepsilon}\to\gamma^t$ in $(\Gamma(\mathbb{T}^d),\overline{\delta})$, we can find (by scaling $\boldsymbol{\eta}$ from $[0,s]$ to $[0,s_\varepsilon]$ and from $[t,1]$ to $[t_\varepsilon,1]$, and using repeatedly the concatenation, see Remark 3.3.2) generalized flows $\boldsymbol{\nu}^\varepsilon$ between γ and η in $[0,1]$, $s_\varepsilon\uparrow s$, $t_\varepsilon\downarrow t$ satisfying:

(a) $\boldsymbol{\nu}^\varepsilon$ connects η to η^s in $[0,s_\varepsilon]$, η^s to $\eta^{s,\varepsilon}$ in $[s_\varepsilon,s]$, $\gamma^{t,\varepsilon}$ to γ^t in $[t,t_\varepsilon]$, γ^t to γ in $[t_\varepsilon,1]$ and is incompressible in all these time intervals;

(b) the restriction of $\boldsymbol{\nu}^\varepsilon$ to $[s,t]$ coincides with $\hat{\boldsymbol{\lambda}}^{\varepsilon}$;

(c) the action of $\boldsymbol{\nu}^\varepsilon$ in $[0,s]$ converges to $\overline{\delta}^2(\eta,\eta^s)=s^2\overline{\delta}^2(\eta,\gamma)$, and the action of $\boldsymbol{\nu}^\varepsilon$ in $[t,1]$ converges to $\overline{\delta}^2(\gamma^t,\gamma)=(1-t)^2\overline{\delta}^2(\eta,\gamma)$.

Since $\boldsymbol{\nu}^\varepsilon$ is a flow with bounded compression connecting η to γ we use (3.6.8) and the incompressibility in $[0,1]\setminus[s,t]$ to obtain

$$\int_{\tilde\Omega(\mathbb{T}^d)}\int_0^1\frac{1}{2}|\dot\omega(\tau)|^2\,d\tau\,d\boldsymbol{\nu}^\varepsilon(\omega,a)-\int_{\tilde\Omega(\mathbb{T}^d)}\int_s^t\bar p(\tau,\omega)\,d\tau\,d\boldsymbol{\nu}^\varepsilon(\omega,a)\geq\overline{\delta}^2(\eta,\gamma) \tag{3.6.23}$$

for all $\varepsilon>0$. Taking into account that (b) and (c) imply

$$\int_{\tilde\Omega(\mathbb{T}^d)}\int_0^s\frac{1}{2}|\dot\omega(\tau)|^2\,d\tau\,d\boldsymbol{\nu}^\varepsilon(\omega,a)\to s\overline{\delta}^2(\eta,\gamma)$$

and

$$\int_{\tilde\Omega(\mathbb{T}^d)}\int_t^1\frac{1}{2}|\dot\omega(\tau)|^2\,d\tau\,d\boldsymbol{\nu}^\varepsilon(\omega,a)\to(1-t)\overline{\delta}^2(\eta,\gamma),$$

from (3.6.22) and (3.6.23) we get

$$2\delta+\int_{\mathbb{T}^d}\Big(W_{c_{\bar{p}}^{s,t}}(\eta^s_a,\gamma^t_a)\vee-\frac{1}{\delta}\Big)\,d\mu_{\mathbb{T}}(a)\geq(1-s-(1-t))\overline{\delta}^2(\eta,\gamma)=(t-s)\overline{\delta}^2(\eta,\gamma). \tag{3.6.24}$$

Letting $\delta\downarrow 0$ we obtain the $\mu_{\mathbb{T}}$-integrability of $W_{c_{\bar{p}}^{s,t}}(\eta^s_a,\gamma^t_a)$ and (3.6.18). □

A byproduct of the above proof is that equalities hold in (3.6.17), (3.6.18), and therefore

$$
\begin{aligned}
&\int_{\mathbb{T}^d}\int_{\Omega(\mathbb{T}^d)}\left(\int_s^t \frac{1}{2}|\dot\omega(\tau)|^2 - \bar p(\tau,\omega)\,d\tau - c_{\bar p}^{s,t}(\omega(s),\omega(t))\right) d\boldsymbol{\eta}_a(\omega)\,d\mu_{\mathbb{T}}(a) \qquad (3.6.25)\\
&= \int_{\mathbb{T}^d}\int_{\Omega(\mathbb{T}^d)}\int_s^t \frac{1}{2}|\dot\omega(\tau)|^2 - \bar p(\tau,\omega)\,d\tau\,d\boldsymbol{\eta}_a(\omega)\,d\mu_{\mathbb{T}}(a) - \int_{\mathbb{T}^d} W_{c_{\bar p}^{s,t}}(\eta_a^s,\gamma_a^t)\,d\mu_{\mathbb{T}}(a) = 0.
\end{aligned}
$$

This yields in particular also the first optimality condition. However, as the proof of Theorem 3.6.11 is much more technical than the one presented in Theorem 3.6.8, we decided to present both.

Now we show that the optimality conditions in Theorems 3.6.8 and 3.6.11 are also sufficient, even in the case of a general compact manifold without boundary D.

Theorem 3.6.12 (Sufficient condition). *Assume that $\boldsymbol{\eta} = \boldsymbol{\eta}_a \otimes \mu$ is a generalized incompressible flow in D between η and γ, and assume that for some map q the following properties hold:*

(a) $Mq \in L^1((0,1)\times D)$ and $\boldsymbol{\eta}$ is concentrated on q-minimizing paths;

(b) the plan $(e_0,e_1)_\#\boldsymbol{\eta}_a$ is optimal, relative to the cost $c_q^{0,1}$ defined in (3.6.13), for μ_D-a.e. a.

Then $\boldsymbol{\eta}$ is optimal and q is the pressure field. In addition, if (a), (b) are replaced by

(a') $Mq \in L^1_{\mathrm{loc}}((0,1)\times D)$ and $\boldsymbol{\eta}$ is concentrated on locally q-minimizing paths;

(b') for all intervals $[s,t]\subset(0,1)$, the plan $(e_s,e_t)_\#\boldsymbol{\eta}_a$ is optimal, relative to the cost $c_q^{s,t}$ defined in (3.6.13), for μ_D-a.e. a,

the same conclusions hold.

Proof. Assume first that (a) and (b) hold, and assume without loss of generality that $\int_D q(t,\cdot)\,d\mu_D = 0$ for almost all $t\in(0,1)$. Recalling that, thanks to the global integrability of Mq, any generalized incompressible flow $\boldsymbol{\nu} = \boldsymbol{\nu}_a \otimes \mu_D$ between η and γ is concentrated on curves ω with $Mq(\tau,\omega)\in L^1(0,1)$ (see Remark 3.6.7), we have

$$
\begin{aligned}
\mathscr{A}_1(\boldsymbol{\nu}) &= \int_D\int_{\Omega(D)}\int_0^1 \frac{1}{2}|\dot\omega|^2 - q(\tau,\omega)\,d\tau\,d\boldsymbol{\nu}_a(\omega)\,d\mu_D(a) \qquad (3.6.26)\\
&\geq \int_D\int_{D\times D} c_q^{0,1}(x,y)\,d(e_0,e_1)_\#\boldsymbol{\nu}_a\,d\mu_D(a) \geq \int_D W_{c_q^{0,1}}(\eta_a,\gamma_a)\,d\mu_D(a).
\end{aligned}
$$

When $\boldsymbol{\nu} = \boldsymbol{\eta}$ the first inequality is an equality, because $\boldsymbol{\eta}$ is concentrated on q-minimizing paths, as well as the second inequality, because of the optimality of the plan $(e_0,e_1)_\#\boldsymbol{\eta}_a$.

This proves that $\boldsymbol{\eta}$ is optimal. Moreover, by using the inequality in (3.6.26) with a flow $\boldsymbol{\nu}$ with bounded compression, one obtains

$$\mathscr{A}_1(\boldsymbol{\nu}) \geq \mathscr{A}_1(\boldsymbol{\eta}) + \langle q, \rho^{\boldsymbol{\nu}} - 1 \rangle.$$

Considering almost incompressible flows $\boldsymbol{\nu}$ arising by a smooth perturbation of $\boldsymbol{\eta}$ as described at the beginning of this section (see (3.6.1) in particular), the same argument used to obtain (3.6.6) gives that q satisfies (3.6.6), so that q is the pressure field.

In the case when (a)' and (b)' hold, by localizing in all intervals $[s,t] \subset (0,1)$ the previous argument (see Remark 3.3.2), one obtains that

$$(t-s) \int_{\tilde{\Omega}(D)} \int_s^t \frac{1}{2} |\dot{\omega}|^2 \, d\tau \, d\boldsymbol{\eta}(\omega, a) = \bar{\delta}^2(\gamma_s, \gamma_t),$$

where $\gamma_s = (e_s, \pi_D)_\# \boldsymbol{\eta}$ and $\gamma_t = (e_t, \pi_D)_\# \boldsymbol{\eta}$. Letting $s \downarrow 0$ and $t \uparrow 1$ we obtain the optimality of $\boldsymbol{\eta}$. ☐

A byproduct of the previous result is a new variational principle satisfied, at least locally in time, by the pressure field. Up to a restriction to a smaller time interval we shall assume that $Mp \in L^1\left([0,1] \times \mathbb{T}^d\right)$.

Corollary 3.6.13 (Variational characterization of the pressure). *Let η, $\gamma \in \Gamma(\mathbb{T}^d)$ and let p be the unique pressure field induced by the constant speed geodesics in $[0,1]$ between $\eta = \eta_a \otimes \mu_{\mathbb{T}}$ and $\gamma = \gamma_a \otimes \mu_{\mathbb{T}}$. Assume that $Mp \in L^1\left([0,1] \times \mathbb{T}^d\right)$ and, with no loss of generality, $\int_{\mathbb{T}^d} p(t, \cdot) \, d\mu_{\mathbb{T}} = 0$. Then $\bar{p}$ maximizes the functional*

$$q \mapsto \Psi(q) := \int_{\mathbb{T}^d} W_{c_q^{0,1}}(\eta_a, \gamma_a) \, d\mu_{\mathbb{T}}(a) + \int_0^1 \int_{\mathbb{T}^d} q(\tau, x) \, d\mu_{\mathbb{T}}(x) \, d\tau$$

among all functions $q : [0,1] \times \mathbb{T}^d \to \mathbb{R}$ with $Mq \in L^1([0,1] \times \mathbb{T}^d)$.

Proof. We first remark that the functional Ψ is invariant under sum of functions depending on t only, so we can assume that the spatial means of any function q vanish. From (3.6.25) we obtain that

$$\int_{\mathbb{T}^d} W_{c_{\bar{p}}^{0,1}}(\eta_a, \gamma_a) \, d\mu_{\mathbb{T}}(a) = \int_{\mathbb{T}^d} \int_{\Omega(\mathbb{T}^d)} \int_0^1 \frac{1}{2} |\dot{\omega}(\tau)|^2 - \bar{p}(\tau, \omega) \, d\tau \, d\boldsymbol{\eta}_a(\omega) \, d\mu_{\mathbb{T}}(a).$$

By the incompressibility constraint, in the right hand side $\bar{p}$ can be replaced by any function q whose spatial means vanish and, if $Mq \in L^1\left([0,1] \times \mathbb{T}^d\right)$, the resulting integral bounds from above $\int_{\mathbb{T}^d} W_{c_q^{0,1}}(\eta_a, \gamma_a) \, d\mu_{\mathbb{T}}(a)$, as we proved in (3.6.26). ☐

3.7 Regularity of the pressure field

In this last section, using the Eulerian-Lagrangian formulation introduced by Brenier in [35], we want to improve his regularity result to deduce that the pressure field is a locally integrable funtion.

We therefore consider the family of distributional solutions $c_{t,a}$, indexed by $a \in D$, of the continuity equation (3.3.13) with the initial and final conditions (3.3.14), and we minimize the action

$$\int_0^T \int_{D\times D} \frac{1}{2}|v|^2(t,x,a)\,dc(t,x,a),$$

under the global constraint given by the incompressibility of the flow (3.3.15). By what we have already proved, the existence of minimizing pairs (c, v) with finite action holds when, for instance, $D = [0,1]^d$ or $D = \mathbb{T}^d$ is the flat d-dimensional torus (see Section 3.3.2). Moreover minimizing pairs (c, v) satisfy the following two properties:

(a) (**Constancy of kinetic energy**) The map $t \mapsto \int |v|^2(t,x,a)\,dc_t$ coincides a.e. in $(0,T)$ with a constant ($2T^{-1}$ times the minimal action);

(b) (**Weak solution to Euler's equations**) There exists a distribution p in $(0,T)\times D$ satisfying

$$\nabla p = -\partial_t \left(\int_D v(t,x,a)\,dc_{t,x}(a) \right) - \mathrm{div} \left(\int_D v(t,x,a) \otimes v(t,x,a)\,dc_{t,x}(a) \right),$$

in the sense of distributions.

In this section we refine a little bit the deep analysis made in [35] of the regularity of the gradient of the pressure field: Brenier proved that the distributions $\partial_{x_i} p$ are locally finite measures in $(0,T)\times D$, but this information is not sufficient (due to a lack of time regularity) to imply that p is a function. As shown in Corollary 3.7.4, a sufficient condition, that gives also $p \in L^2_{\mathrm{loc}}\big((0,T); L^{d/(d-1)}_{\mathrm{loc}}(D)\big)$, is that

$$\partial_{x_i} p \in L^2_{\mathrm{loc}}\big((0,T); \mathcal{M}_{\mathrm{loc}}(D)\big), \qquad i = 1, \ldots, d.$$

The proof of this regularity property is the main scope of this section. The fact that p is a function at least in some $L^1_{\mathrm{loc}}(L^r_{\mathrm{loc}})$ space, for some $r > 1$, plays an important role in the analysis, developed in Section 3.6, of the necessary and sufficient optimality conditions for action-minimizing curves in $\Gamma(D)$. Indeed, these conditions involve the Lagrangian

$$\mathcal{L}_p(\gamma) := \int \frac{1}{2}|\dot\gamma(t)|^2 - p(t,\gamma(t))\,dt,$$

the (locally) minimizing curves for $\mathcal{L}_p$ and the value function induced by $\mathcal{L}_p$, and none of these objects makes sense if p is only a measure in the time variable.

From now, we fix a minimizing pair (c, v), and we shall denote by

$$A^* := \frac{1}{2}\int_0^T \int_{D\times D} |v|^2(t,x,a)dc(t,x,a) = \frac{1}{2}T\int_{D\times D}\int |v|^2(t,x,a)\,dc_t(x,a)$$

its action (the last equality follows from the property (a) stated above). To simplify our notation we just denote by $\int$ the integration on the whole space $(0,T)\times D\times D$, whenever no ambiguity arises. We shall also assume that either D is the closure of a bounded Lipschitz domain in $\mathbb{R}^d$, or that $D = \mathbb{T}^d$ is the d-dimensional flat torus, and denote by $\int dx$ the integration with respect to μ_D.

3.7.1 A difference quotients estimate

In order to proceed to the proof, we recall an approximation of the pressure field obtained in [35] through a dual formulation. The arguments in [35] extend with no change to the more general model described in the introduction, where an initial and final measure-preserving plan (instead of $\boldsymbol{i}$ and a measure-preserving map f) are considered.

Let us consider the Banach space $E := C^0(\tilde{Q}) \times [C^0(\tilde{Q})]^d$, where $\tilde{Q} := [0,T]\times D\times D$, and we define the convex functions $\alpha : E \to (-\infty,\infty]$ and $\beta : E \to (-\infty,\infty]$ given by

$$\alpha(F,\Phi) := \begin{cases} 0 & \text{if } F + \frac{1}{2}|\Phi|^2 \le 0, \\ +\infty & \text{otherwise,} \end{cases}$$

$$\beta(F,\Phi) := \begin{cases} \langle c, F\rangle + \langle vc, \Phi\rangle & \text{if } F = -\partial_t\phi - p,\ \Phi = -\nabla_x\phi, \\ & \text{for some } \phi\in C^0(\tilde{Q}) \text{ and } p \in C^0([0,T]\times D), \\ +\infty & \text{otherwise,} \end{cases}$$

where (c, v) is the fixed minimizing pair. By the Fenchel-Rockafeller duality Theorem, Brenier proved in [35, Section 3.2] that

$$\sup_{(F,\Phi)\in E}\{-\alpha(-F,-\Phi) - \beta(F,\Phi)\} = \inf_{(\tilde{c},\tilde{v}\tilde{c})\in E^*}\{\alpha^*(\tilde{c},\tilde{v}\tilde{c}) + \beta^*(\tilde{c},\tilde{v}\tilde{c})\},$$

where α^* and β^* denote the Legendre-Fenchel transforms of α and β respectively. Writing explicitly the minimization problem appearing in the right hand side, one exactly recovers the minimization of the action $\frac{1}{2}\int |v|^2\,dc$, coupled with the endpoint and incompressibility constraints (3.3.14) and (3.3.15). Indeed

$$\alpha^*(\tilde{c},\tilde{v}\tilde{c}) = \frac{1}{2}\langle|\tilde{v}|^2,\tilde{c}\rangle = \frac{1}{2}\int |\tilde{v}|^2\,d\tilde{c},$$

and

$$\beta^*(\tilde{c}, \tilde{v}\tilde{c}) := \begin{cases} 0 & \text{if } \langle c - \tilde{c}, \partial_t \phi + p\rangle + \langle vc - \tilde{v}\tilde{c}, \nabla_x \phi\rangle = 0 \quad \forall\, p, \phi, \\ +\infty & \text{otherwise.} \end{cases}$$

Thus it is simple to check that $\beta^*(\tilde{c}, \tilde{v}\tilde{c}) = 0$ if and only if the two constraints (3.3.14) and (3.3.15) are satisfied.

One therefore deduces that the minimum of the action coincides with the dual problem $\sup_{(F,\Phi)\in E}\{-\alpha(-F, -\Phi) - \beta(F, \Phi)\}$, which more concretely can be written as

$$\sup_{p,\phi} \langle c, \partial_t \phi + p\rangle + \langle vc, \nabla_x \phi\rangle,$$

with

$$\partial_t \phi + \frac{1}{2}|\nabla_x \phi|^2 + p \leq 0.$$

Thus, the duality tells us that, for any $\varepsilon > 0$, there exist $p_\varepsilon(t, x)$ and $\phi_\varepsilon(t, x, a)$ satisfying

$$\partial_t \phi_\varepsilon + \frac{1}{2}|\nabla_x \phi_\varepsilon|^2 + p_\varepsilon \leq 0$$

and

$$\frac{1}{2}\langle |v|^2, c\rangle \leq \langle c, \partial_t \phi_\varepsilon + p_\varepsilon\rangle + \langle vc, \nabla_x \phi_\varepsilon\rangle + \varepsilon^2.$$

As shown in [35, Section 3.2], from this one deduces the estimate

$$\frac{1}{2}\int |v - \nabla_x \phi_\varepsilon|^2 \, dc \leq \varepsilon^2. \tag{3.7.1}$$

We remark that, up to adding to ϕ_ε a function of time, one can always assume $\int_D p_\varepsilon(t, x)\, dx = 0$ for all $t \in [0, T]$. As shown in [35, Section 3.4], the family p_ε in compact in the sense of distributions, so that there exists a cluster point p. Moreover, since any limit point p of p_ε is seen to satisfy (3.6.7) in the sense of distribution for any minimizing pair (c, v), ∇p is uniquely determined, and this enforces the convergence of the whole family $(\nabla p_\varepsilon)_{\varepsilon>0}$ to ∇p in the sense of distributions.

Let us now prove the following regularity result on $\nabla_x \phi_\varepsilon$: we present a proof slightly different from the one in [35].

Proposition 3.7.1. *Let $\tau \in (0, T)$, let $w : \overline{D} \to \mathbb{R}^d$ be a smooth divergence-free vector field parallel to ∂D and let $e^{sw}(x)$ be the measure-preserving flow in D generated by w. Then, for $\eta < \tau$ we have*

$$\int_\tau^{T-\tau} \int_{D\times D} \left|\nabla_x \phi_\varepsilon(t + \eta, e^{\delta w}(x), a) - \nabla_x \phi_\varepsilon(t, x, a)\right|^2 dc \leq L(\varepsilon^2 + \eta^2 + \delta^2), \tag{3.7.2}$$

with L depending only on τ, w, T and A^*.

Proof. In the sequel we fix a cut-off function $\zeta : [0,T] \to [0,1]$ identically equal to 1 on $[\tau, T-\tau]$. We recall the following estimate (Proposition 3.1 in [35]), which follows by the "quasi optimality" of $(p_\varepsilon, \phi_\varepsilon)$ in the dual problem:

$$\frac{1}{2}\int \left|(\partial_t + v^\eta \cdot \nabla_x) e^{\delta\zeta w} - \nabla_x \phi_\varepsilon \circ e^{\delta\zeta w}\right|^2 \, dc^\eta \\ \leq \varepsilon^2 + \frac{1}{2}\int \left|(\partial_t + v^\eta \cdot \nabla_x) e^{\delta\zeta w}\right|^2 \, dc^\eta - \frac{1}{2}\int |v|^2 \, dc, \quad (3.7.3)$$

(here $e^{\delta\zeta w}(x)$ is the flow generated by w starting from x, at time $\delta\zeta$) where (v^η, c^η) is the "reparameterization" of (v,c) given by

$$c^\eta = c^\eta(t)dt = c_{t+\eta\zeta(t)}dt, \qquad v^\eta(t,x,a) = (1+\eta\zeta'(t))v(t+\eta\zeta(t),x,a).$$

The minimality of (v,c) gives $\int |v^\eta|^2 \, dc^\eta \geq \int |v|^2 \, dc$, and the constancy of $t \mapsto \int |v|^2(t,x,a) \, dc_t$ gives

$$\int |v^\eta|^2 \, dc^\eta - \int |v|^2 \, dc = \int (\eta^2(\zeta')^2 + 2\eta\zeta') \, dc \leq C\eta^2, \quad (3.7.4)$$

with C depending on T, A^*, and ζ.

Since c is a weak solution to the incompressible Euler equations and w is divergence-free, we have

$$\int v \cdot (\partial_t + v \cdot \nabla_x)(\zeta w) \, dc = 0.$$

As a consequence, performing a change of variable in time, it is simple to check that

$$\int v^\eta \cdot (\partial_t + v^\eta \cdot \nabla_x)(\zeta w) \, dc^\eta = O(\eta). \quad (3.7.5)$$

If we now add and subtract v^η, we can rewrite (3.7.3) as

$$\int \left|(\partial_t + v^\eta \cdot \nabla_x)(e^{\delta\zeta w}(x) - x) + (v^\eta - \nabla_x \phi_\varepsilon \circ e^{\delta\zeta w})\right|^2 \, dc^\eta \\ \leq 2\varepsilon^2 + \int \left|(\partial_t + v^\eta \cdot \nabla_x)(e^{\delta\zeta w}(x) - x) + v^\eta\right|^2 \, dc^\eta - \int |v|^2 \, dc.$$

Rearranging the squares we get

$$\int \left|v^\eta - \nabla_x \phi_\varepsilon \circ e^{\delta\zeta w}\right|^2 \, dc^\eta \leq -2\int \left[(\partial_t + v^\eta \cdot \nabla_x)(e^{\delta\zeta w}(x) - x)\right] \cdot \left[v^\eta - \nabla_x \phi_\varepsilon \circ e^{\delta\zeta w}\right] \, dc^\eta \\ + 2\varepsilon^2 + 2\int v^\eta \cdot (\partial_t + v^\eta \cdot \nabla_x)(e^{\delta\zeta w}(x) - x) \, dc^\eta \\ + \int |v^\eta|^2 \, dc^\eta - \int |v|^2 \, dc.$$

Defining

$$f(\delta,\varepsilon,\eta) := \int \left|v^\eta - \nabla_x\phi_\varepsilon \circ e^{\delta\zeta w}\right|^2 dc^\eta = \int \left|(1+\eta\zeta')v(1+\eta\zeta,x,a) - \nabla_x\phi_\varepsilon(1+\eta\zeta, e^{\delta\zeta w}(x),a)\right|^2 dc$$
$$\geq \int_\tau^{T-\tau}\int_{D\times D}\left|v - \nabla_x\phi_\varepsilon(t+\eta, e^{\delta w}(x),a)\right|^2 dc$$

we see that it suffices to bound f from above. Since $e^{\delta\zeta w}x - x = \delta\zeta(t)w(x) + O(\delta^2)$ (in the C^1 norm in spacetime), by Schwarz inequality, (3.7.4) and (3.7.5) we get

$$f \leq C\sqrt{f}\delta + 2\varepsilon^2 + C(\delta\eta+\delta^2) + C\eta^2,$$

which implies $f(\delta,\varepsilon,\eta) \leq C(\delta^2+\varepsilon^2+\eta^2)$, with C depending on T, A^*, ζ, and w. This, together with $\int |v-\nabla_x\phi_\varepsilon|^2\, dc \leq 2\varepsilon^2$, gives (3.7.2). □

3.7.2 Proof of the main result

Theorem 3.7.2. *Let $\tau \in (0,T)$ and let $w : \overline{D} \to \mathbb{R}^d$ be a smooth divergence-free vector field parallel to ∂D. Then there exists a constant $C = C(w,\tau,T,A^*)$ such that*

$$|\langle \nabla p\cdot w, \zeta f\rangle| \leq C\|f\|_\infty \|\zeta\|_{L^2(0,T)} \qquad \forall \zeta \in C^\infty_c\big((\tau,T-\tau);[0,+\infty)\big),\ f \in C^\infty_c\big((0,1)\times D\big). \tag{3.7.6}$$

Proof. For $\zeta \in C^\infty_c(\tau,T-\tau)$ nonnegative, $\eta \in (0,\tau/2)$ and δ, $\varepsilon > 0$ we consider the following expression:

$$I = I(\zeta,\delta,\eta,\varepsilon) := \int_0^T\int_D \zeta(t)\left|\int_0^1 \left[p_\varepsilon(t+\eta\theta, e^{\delta w}(x)) - p_\varepsilon(t+\eta\theta,x)\right] d\theta\right| dxdt$$
$$= \int \zeta(t)\left|\int_0^1 \left[p_\varepsilon(t+\eta\theta, e^{\delta w}(x)) - p_\varepsilon(t+\eta\theta,x)\right] d\theta\right| dc(t,x,a).$$

Our goal is to bound I from above. This will be achieved in the following (many) steps: $I \leq I_1+I_2+I_3$ and estimate of I_2, I_3; $I_1 \leq 2\|\zeta\|_\infty\varepsilon^2 - (I_4+I_5+I_6)$ and estimate of I_5 and I_6; $I_4 = I_7+I_8$ and estimate of I_8; $I_7 = 2I_9+I_{10}$ and estimate of I_9; $I_{10} = I_{11}+I_{12}$ and estimate of I_{12}; finally $I_{11} = I_{13}+I_{14}$ and estimate of I_{13} and I_{14}. In order to avoid a cumbersome notation, during this proof we denote by C a generic constant depending only on (w,τ,T,A^*), whose specific value can change from line to line.

We now consider $\lambda_\varepsilon(t,x,a) := -\big(\partial_t\phi_\varepsilon + \frac{1}{2}|\nabla_x\phi_\varepsilon|^2 + p_\varepsilon\big) \geq 0$, and we recall that $\int \lambda_\varepsilon\, dc \leq \varepsilon^2$. We have

$$I \leq I_1+I_2+I_3,$$

where

$$I_1 := \int \zeta(t) \left| \int_0^1 \left[\lambda_\varepsilon(t+\eta\theta, e^{\delta w}(x), a) - \lambda_\varepsilon(t+\eta\theta, x, a) \right] d\theta \right| dc,$$

$$I_2 := \int \zeta(t) \left| \int_0^1 \left[\partial_t \phi_\varepsilon(t+\eta\theta, e^{\delta w}(x), a) - \partial_t \phi_\varepsilon(t+\eta\theta, x, a) \right] d\theta \right| dc,$$

$$I_3 := \int \zeta(t) \left| \int_0^1 \left[\frac{1}{2} |\nabla_x \phi_\varepsilon|^2 (t+\eta\theta, e^{\delta w}(x), a) - \frac{1}{2} |\nabla_x \phi_\varepsilon|^2 (t+\eta\theta, x, a) \right] d\theta \right| dc.$$

By (3.7.2) we have

$$\|\nabla_x \phi_\varepsilon(t+\eta\theta, e^{\delta w}(x), a)\|_{L^2(\zeta^2 c)} \leq \|\nabla_x \phi_\varepsilon(t, x, a)\|_{L^2(\zeta^2 c)} + \sqrt{L} \|\zeta\|_\infty (\varepsilon + \eta + \delta) \qquad \forall \theta \in (0,1).$$

Therefore writing $|A|^2 - |B|^2$ as $(A-B)\cdot(A+B)$ and using (3.7.2) once more, we can estimate

$$I_3 \leq C(\varepsilon+\eta+\delta) \left(\int \zeta^2(t) |\nabla_x \phi_\varepsilon|^2 (t, x, a)\, dc + C \|\zeta\|_\infty^2 (\varepsilon^2 + \eta^2 + \delta^2) \right)^{1/2}. \tag{3.7.7}$$

For I_2 we first integrate with respect to θ and then use the mean value theorem to obtain

$$\begin{aligned} I_2 &\leq \frac{\delta}{\eta} \int \zeta(t) \int_0^1 \left| \left[\nabla_x \phi_\varepsilon(t+\eta, e^{\sigma\delta w}(x), a) - \nabla_x \phi_\varepsilon(t, e^{\sigma\delta w}(x), a) \right] \cdot w(e^{\sigma\delta w}(x)) \right| d\sigma dc \\ &\leq C \frac{\delta}{\eta} \int_0^1 \int \zeta(t) \left| \left[\nabla_x \phi_\varepsilon(t+\eta, e^{\sigma\delta w}(x), a) - \nabla_x \phi_\varepsilon(t, e^{\sigma\delta w}(x), a) \right| dc d\sigma \\ &\leq C \frac{\delta}{\eta} (\varepsilon+\eta+\delta) \|\zeta\|_{L^2(0,T)}. \end{aligned} \tag{3.7.8}$$

Let us now consider I_1: using $\lambda_\varepsilon \geq 0$ and $\int \lambda_\varepsilon \, dc \leq \varepsilon^2$, we obtain

$$\begin{aligned} I_1 &\leq \int \zeta(t) \int_0^1 \left[\lambda_\varepsilon(t+\eta\theta, e^{\delta w}(x), a) + \lambda_\varepsilon(t+\eta\theta, x, a) \right] d\theta dc \\ &\leq 2\|\zeta\|_\infty \varepsilon^2 + \int \zeta(t) \int_0^1 \left[\lambda_\varepsilon(t+\eta\theta, e^{\delta w}(x), a) + \lambda_\varepsilon(t+\eta\theta, x, a) - 2\lambda_\varepsilon(t, x, a) \right] d\theta dc \\ &\leq 2\|\zeta\|_\infty \varepsilon^2 - I_4 - I_5 - I_6, \end{aligned}$$

where

$$I_4 := \int \zeta(t) \int_0^1 \left[\partial_t \phi_\varepsilon(t+\eta\theta, e^{\delta w}(x), a) + \partial_t \phi_\varepsilon(t+\eta\theta, x, a) - 2\partial_t \phi_\varepsilon(t, x, a) \right] d\theta dc,$$

$$I_5 := \frac{1}{2} \int \zeta(t) \int_0^1 \left[|\nabla_x \phi_\varepsilon|^2 (t+\eta\theta, e^{\delta w}(x), a) + |\nabla_x \phi_\varepsilon|^2 (t+\eta\theta, x, a) - 2|\nabla_x \phi_\varepsilon|^2 (t, x, a) \right] d\theta dc,$$

$$I_6 := \int \zeta(t) \int_0^1 \left[p_\varepsilon(t+\eta\theta, e^{\delta w}(x)) + p_\varepsilon(t+\eta\theta, x) - 2p_\varepsilon(t, x) \right] d\theta dc.$$

Now we notice that

$$I_6 = 0, \tag{3.7.9}$$

since $\int c(t,x,da) = 1$ (by the incompressibility constraint (3.3.15)), $e^{\delta w}$ is measure-preserving, and $\int p_\varepsilon(t,x)\,dx = 0$. For I_5, we have the same bound as for I_3, that is

$$|I_5| \le C(\varepsilon+\eta+\delta)\left(\int \zeta^2(t)|\nabla_x\phi_\varepsilon|^2(t,x,a)\,dc + C\|\zeta\|_\infty^2(\varepsilon^2+\eta^2+\delta^2)\right)^{1/2}. \tag{3.7.10}$$

We continue splitting I_4 as $I_7 + I_8$, with

$$\begin{aligned}
I_7 :&= \int\int_0^1 \big[\zeta(t)\big(\partial_t\phi_\varepsilon(t+\eta\theta,e^{\delta w}(x),a) + \partial_t\phi_\varepsilon(t+\eta\theta,x,a)\big) - 2\zeta(t-\theta\eta)\partial_t\phi_\varepsilon(t,x,a)\big]\,d\theta dc,\\
I_8 :&= 2\int\int_0^1 \big[\zeta(t-\theta\eta)-\zeta(t)\big]\partial_t\phi_\varepsilon(t,x,a)\,d\theta dc.
\end{aligned}$$

For I_8, using once more that $\lambda_\varepsilon \ge 0$ and $\int \lambda_\varepsilon\,dc \le \varepsilon^2$, we have the bound

$$\begin{aligned}
|I_8| &\le 2\left|\int\int_0^1 \big[\zeta(t-\theta\eta)-\zeta(t)\big]\lambda_\varepsilon(t,x,a)\,d\theta dc\right| + \left|\int\int_0^1 \big[\zeta(t-\theta\eta)-\zeta(t)\big]|\nabla_x\phi_\varepsilon|^2(t,x,a)\,d\theta dc\right|\\
&\quad + 2\left|\int\int_0^1 \big[\zeta(t-\theta\eta)-\zeta(t)\big]p_\varepsilon(t,x,a)\,d\theta dc\right|\\
&\le 4\|\zeta\|_\infty\varepsilon^2 + \left|\int\int_0^1 \big[\zeta(t-\theta\eta)-\zeta(t)\big]|\nabla_x\phi_\varepsilon|^2(t,x,a)\,d\theta dc\right|
\end{aligned}$$

where in the last inequality we used that $\int p_\varepsilon\,dc_t = \int p_\varepsilon\,dx = 0$. Using also the fact that $t \mapsto \int |v|^2(t,x,a)\,dc_t$ does not depend on t we get

$$|I_8| \le 4\|\zeta\|_\infty\varepsilon^2 + 2\|\zeta\|_\infty \int \big||\nabla_x\phi_\varepsilon|^2(t,x,a) - |v|^2(t,x,a)\big|\,dc. \tag{3.7.11}$$

We now consider $I_7 = I_9 + 2I_{10}$, where

$$\begin{aligned}
I_9 :&= \int \zeta(t)\int_0^1 \big[\partial_t\phi_\varepsilon(t+\eta\theta,e^{\delta w}(x),a) - \partial_t\phi_\varepsilon(t+\eta\theta,x,a)\big]\,d\theta dc,\\
I_{10} :&= \int\int_0^1 \big[\zeta(t)\partial_t\phi_\varepsilon(t+\eta\theta,x,a) - \zeta(t-\theta\eta)\partial_t\phi_\varepsilon(t,x,a)\big]\,d\theta dc.
\end{aligned}$$

We have, as for I_2,

$$
\begin{aligned}
|I_9| &= \frac{1}{\eta}\left|\int \zeta(t)\big[\big(\phi_\varepsilon(t+\eta, e^{\delta w}(x), a) - \phi_\varepsilon(t+\eta, x, a)\big) - \big(\phi_\varepsilon(t, e^{\delta w}(x), a) - \phi_\varepsilon(t, x, a)\big)\big]\, dc\right| \\
&= \frac{\delta}{\eta}\left|\int \zeta(t)\int_0^1 \big[\nabla_x\phi_\varepsilon(t+\eta, e^{\sigma\delta w}(x), a) - \nabla_x\phi_\varepsilon(t, e^{\sigma\delta w}(x), a)\big]\cdot w(e^{\sigma\delta w}(x))\, d\sigma dc\right| \\
&\leq C\frac{\delta}{\eta}(\varepsilon+\eta+\delta)\|\zeta\|_{L^2(0,T)}.
\end{aligned}
\tag{3.7.12}
$$

For I_{10}, we use the continuity equation $\partial_t c + \operatorname{div}_x(vc) = 0$ (see (3.3.16)) and add and subtract $\zeta(t)$ to get

$$
\begin{aligned}
I_{10} &= \int\int_0^1\int_0^1 \partial_t\big[\zeta(t-(1-\sigma)\eta\theta)\partial_t\phi_\varepsilon(t+\eta\theta\sigma, x, a)\big]\eta\theta\, d\sigma d\theta dc \\
&= -\int\int_0^1\int_0^1 \zeta(t-(1-\sigma)\eta\theta)\partial_t\nabla_x\phi_\varepsilon(t+\eta\theta\sigma, x, a)\cdot v(t,x,a)\eta\theta\, d\sigma d\theta dc \\
&= -\int\int_0^1\int_0^1 \big[\zeta(t-(1-\sigma)\eta\theta) - \zeta(t)\big]\partial_t\nabla_x\phi_\varepsilon(t+\eta\theta\sigma, x, a)\cdot v(t,x,a)\eta\theta\, d\sigma d\theta dc \\
&- \int\int_0^1\int_0^1 \zeta(t)\partial_t\nabla_x\phi_\varepsilon(t+\eta\theta\sigma, x, a)\cdot v(t,x,a)\eta\theta\, d\sigma d\theta dc \\
&= -\int\int_0^1\int_0^1 \big[\zeta(t-(1-\sigma)\eta\theta) - \zeta(t)\big]\partial_t\nabla_x\phi_\varepsilon(t+\eta\theta\sigma, x, a)\cdot v(t,x,a)\eta\theta\, d\sigma d\theta dc \\
&- \int \zeta(t)\int_0^1 \big[\nabla_x\phi_\varepsilon(t+\eta\theta, x, a) - \nabla_x\phi_\varepsilon(t, x, a)\big]\cdot v(t,x,a)\, d\theta dc \\
&=: I_{11} + I_{12}.
\end{aligned}
$$

Now we see that, using (3.7.2) and the Schwarz inequality, we easily get

$$
|I_{12}| \leq C(\varepsilon+\eta)\left(\int \zeta^2(t)|\nabla_x\phi_\varepsilon|^2\, dc + C\|\zeta\|_\infty^2(\varepsilon^2+\eta^2)\right)^{\frac{1}{2}}. \tag{3.7.13}
$$

For I_{11}, it can be written as $I_{13}+I_{14}$, where

$$\begin{aligned}
I_{13} &:= \int\int_0^1\int_0^1 \partial_t\big[[\zeta(t-(1-\sigma)\eta\theta)-\zeta(t)]\nabla_x\phi_\varepsilon(t+\eta\theta\sigma,x,a)\big]\cdot v(t,x,a)\eta\theta\,d\sigma d\theta dc\\
&= \int\int_0^1 [\zeta(t-\eta\theta)-\zeta(t)]\nabla_x\phi_\varepsilon(t,x,a)\cdot v(t,x,a)\,d\theta dc,\\
&= \int\int_0^1 [\zeta(t-\eta\theta)-\zeta(t)]\big[\nabla_x\phi_\varepsilon(t,x,a)-v(t,x,a)\big]\cdot v(t,x,a)\,d\theta dc\\
&- \int\int_0^1 [\zeta(t-\eta\theta)-\zeta(t)]|v|^2(t,x,a)\,d\theta dc
\end{aligned}$$

and

$$I_{14} := \int\int_0^1\int_0^1 \big[\zeta'(t-(1-\sigma)\eta\theta)-\zeta'(t)\big]\nabla_x\phi_\varepsilon(t+\eta\theta\sigma,x,a)\cdot v(t,x,a)\eta\theta\,d\sigma d\theta dc.$$

Recalling that $t\mapsto\int|v|^2(t,x,a)\,dc_t$ is constant, by (3.7.1) we have

$$|I_{13}|\le\left|\int\int_0^1[\zeta(t-\eta\theta)-\zeta(t)]\big(\nabla_x\phi_\varepsilon(t,x,a)-v(t,x,a)\big)\cdot v(t,x,a)\,d\theta dc\right|\le C\|\zeta\|_\infty\varepsilon. \tag{3.7.14}$$

Finally, by (3.7.2) we can bound I_{14} with

$$\begin{aligned}
|I_{14}| &\le \|\zeta''\|_\infty\eta^2\int_{\tau/2}^{T-\tau/2}\int_{D\times A}\int_0^1\int_0^1\big|\nabla_x\phi_\varepsilon(t+\eta\theta\sigma,x,a)\cdot v(t,x,a)\big|\,d\sigma d\theta dc\\
&\le \|\zeta''\|_\infty\eta^2 C\left(\|\nabla_x\phi_\varepsilon\|_2+C(\varepsilon+\eta)\right). \qquad (3.7.15)
\end{aligned}$$

Collecting (3.7.7), (3.7.8), (3.7.9), (3.7.10), (3.7.12), (3.7.13), (3.7.14), (3.7.15) we can bound from above I as follows:

$$\begin{aligned}
&C(\varepsilon+\eta+\delta)\left(\int\zeta^2(t)|\nabla_x\phi_\varepsilon|^2\,dc+C\|\zeta\|_\infty^2(\varepsilon^2+\eta^2+\delta^2)\right)^{\frac12}\\
&+I_8+C\frac{\delta}{\eta}(\varepsilon+\eta+\delta)\|\zeta\|_{L^2(0,T)}+\|\zeta''\|_\infty\eta^2C\left(\|\nabla_x\phi_\varepsilon\|_2+C(\varepsilon+\eta)\right)+2\|\zeta\|_\infty\varepsilon^2+C\|\zeta\|_\infty\varepsilon.
\end{aligned}$$

Now, recalling the definition of I, we integrate $p_\varepsilon\zeta$ against a function $f\in C_c^\infty\big((0,T)\times D\big)$ and pass to the limit as $\varepsilon\to 0$, with $\eta=\delta$ frozen, to obtain

$$\frac{1}{\delta}\left|\int_0^1\langle q,\zeta(t)\big[f(t-\delta\theta,e^{-\delta w}(x))-f(t-\delta\theta,x)\big]\rangle\,d\theta\right|\le C\|f\|_\infty(\|\zeta\|_{L^2(0,T)}+\delta\|\zeta''\|_\infty+\delta\|\zeta\|_\infty)$$

for any limit point q of p_ε in the sense of distributions, thanks to the fact that, by (3.7.11), $I_8 \to 0$ as $\varepsilon \to 0$ (here we use again that $t \mapsto \int |v|^2(t,x,a)\,dc_t$ is constant). So, letting $\delta \to 0$, we finally obtain (3.7.6), with ∇q in place of ∇p. But $\nabla p_\varepsilon \to \nabla p$ implies that $\nabla p = \nabla q$ and concludes the proof. $\square$

Remark 3.7.3. In the case $D = \mathbb{T}^d$ one can also consider constant vector fields w and therefore (3.7.6) holds in a stronger (and simpler) form:

$$|\langle \partial_{x_i} p, \zeta f\rangle| \le C\|f\|_\infty \|\zeta\|_{L^2(0,T)} \qquad \forall \zeta \in C_c^\infty\big((\tau, T-\tau); [0,+\infty)\big),\ f \in C^\infty\big([0,1]\times\mathbb{T}^d\big) \tag{3.7.16}$$

with C depending only on τ, T and A^*.

A simple localization and smoothing argument based on (3.7.6) gives that the pressure field is locally (globally, in the case $D = \mathbb{T}^d$) induced by a function.

Corollary 3.7.4. *Let $d \ge 2$. Then for all smooth subdomains $D' \subset\subset D$ there exists*

$$q \in L^2_{\mathrm{loc}}\big((0,T); BV(D')\big) \subset L^2_{\mathrm{loc}}\big((0,T); L^{1^*}(D')\big)$$

(here $1^* = d/(d-1)$) with $\nabla q = \nabla p$ in the sense of distributions in $(0,T)\times D'$. In the case $D = \mathbb{T}^d$ the same statement holds globally in space, i.e. with $D' = D$. Moreover, in this case the result holds also for $d = 1$ (with $1^* = \infty$).

Proof. We first notice that for $d \ge 2$ any constant vector field $\bar w$ in D' can be extended to a divergence-free, smooth and compactly supported vector field in D: indeed, if $D' \subset\subset D_1 \subset\subset D_2 \subset\subset D$, with D_1 and D_2 smooth, we may set $\hat w = \bar w$ in D_1, $\hat w = 0$ in $D \setminus D_2$, and $\hat w = \nabla\psi$ in $D_2 \setminus \overline{D}_1$, where ψ is a solution of

$$\begin{cases} \Delta\psi = 0 & \text{in } D_2 \setminus \overline{D}_1, \\ \frac{\partial\psi}{\partial\nu} = 0 & \text{on } \partial D_2, \\ \frac{\partial\psi}{\partial\nu} = \bar w\cdot\nu & \text{on } \partial D_1, \end{cases}$$

(existence of ψ can be obtained by minimizing $\frac12\int_{D_2\setminus\overline{D}_1}|\nabla\phi|^2 - \int_{\partial D_1}\phi\bar w\cdot\nu$ in $H^1(D_2\setminus\overline{D}_1)$). By construction $\hat w$ is divergence-free (in the sense of distributions) in D, compactly supported and coincides with $\bar w$ in a neighbourhood of $\overline{D'}$, so that a suitable mollification of $\hat w$ provides the required extension.

Thanks to this remark, (3.7.6) yields

$$|\langle \partial_{x_i} p, \zeta f\rangle| \le L\|f\|_\infty \|\zeta\|_{L^2(0,T)} \qquad \forall \zeta \in C_c^\infty\big((\tau, T-\tau); [0,+\infty)\big),\ f \in C_c^\infty\big((0,1)\times D'\big), \tag{3.7.17}$$

with L depending only on τ, T, D' and A^*. If we denote by q_ε the mollified functions of p, this easily implies that $|\nabla q_\varepsilon|$ is uniformly bounded in $L^2_{\mathrm{loc}}\big((0,T); L^1(D')\big)$. In particular, if we denote by $\bar{q}_\varepsilon$ the mean value of q_ε on D', $q_\varepsilon - \bar{q}_\varepsilon$ is uniformly bounded in the space $L^2_{\mathrm{loc}}\big((0,T);L^{1^*}(D')\big)$, and if q is any weak limit point (in the duality with $L^2_{\mathrm{loc}}\big((0,T);L^d(D')\big)$) we easily get $\nabla q = \nabla p$ and $q \in L^2_{\mathrm{loc}}\big((0,T);BV(D')\big)$.

In the case $D = \mathbb{T}^d$ the proof is analogous: it suffices to apply Remark 3.7.3. $\square$

Chapter 4

On the structure of the Aubry set and Hamilton-Jacobi equation

4.1 Introduction

[1] Let M be a smooth manifold without boundary. We denote by TM the tangent bundle and by $\pi : TM \to M$ the canonical projection. A point in TM will be denoted by (x, v) with $x \in M$ and $v \in T_xM = \pi^{-1}(x)$. In the same way a point of the cotangent bundle T^*M will be denoted by (x, p) with $x \in M$ and $p \in T_x^*M$ a linear form on the vector space T_xM. We will suppose that g is a complete Riemannian metric on M. For $v \in T_xM$, the norm $\|v\|_x$ is $g_x(v, v)^{1/2}$. We will denote by $\|\cdot\|_x$ the dual norm on T^*M. Moreover, for every pair $x, y \in M$, $d(x, y)$ will denote the Riemannian distance from x to y.

We will assume in the whole chapter that $H : T^*M \to \mathbb{R}$ is a Hamiltonian of class $C^{k,\alpha}$, with $k \geq 2, \alpha \in [0, 1]$, which satisfies the three following conditions:

(H1) **C^2-strict convexity:** $\forall (x, p) \in T^*M$, the second derivative along the fibers $\frac{\partial^2 H(x,p)}{\partial p^2}$ is positive strictly definite;

(H2) **uniform superlinearity:** for every $K \geq 0$ there exists a finite constant $C(K)$ such that

$$\forall (x, p) \in T^*M, \quad H(x, p) \geq K\|p\|_x + C(K);$$

(H3) **uniform boundedness in the fibers:** for every $R \geq 0$, we have

$$\sup_{x \in M} \{H(x, p) \mid \|p\|_x \leq R\} < +\infty.$$

[1] This chapter is based on a joint work with Albert Fathi and Ludovic Rifford [67].

By the Weak KAM Theorem we know that, under the above conditions, there is $c(H) \in \mathbb{R}$ such that the Hamilton-Jacobi equation

$$H(x, d_x u) = c$$

admits a global viscosity solution $u : M \to \mathbb{R}$ for $c = c(H)$ and does not admit such solution for $c < c(H)$, see [62], [52], [65], [68], [96]. In fact, if M is assumed to be compact, then $c(H)$ is the only value of c for which the Hamilton-Jacobi equation above admits a viscosity solution. The constant $c(H)$ is called the *critical value*, or the *Mañé critical value* of H. In the sequel, a viscosity solution $u : M \to \mathbb{R}$ of $H(x, d_x u) = c(H)$ will be called a *critical viscosity solution* or a *weak KAM solution*, while a viscosity subsolution u of $H(x, d_x u) = c(H)$ will be called a *critical viscosity subsolution* (or *critical subsolution* if u is at least C^1).

We recall that the Lagrangian $L : TM \to \mathbb{R}$ associated to the Hamiltonian H is defined by

$$\forall (x, v) \in TM, \quad L(x, v) := \max_{p \in T_x^* M} \left\{ p(v) - H(x, p) \right\}.$$

Since H is of class at least C^2 and satisfies the three conditions (H1)-(H3), it is well-known (see for instance [65] or [68, Lemma 2.1])) that L is finite everywhere of class at least C^2, strictly convex and superlinear in each fiber $T_x M$, and satisfies

$$\forall (x, p) \in T_x^* M, \quad H(x, p) = \max_{v \in T_x M} \left\{ p(v) - L(x, v) \right\}.$$

Therefore the Fenchel inequality is always satisfied

$$p(v) \leq L(x, v) + H(x, p).$$

Moreover, we have equality in the Fenchel inequality if and only if

$$(x, p) = \mathcal{L}(x, v),$$

where $\mathcal{L} : TM \to T^* M$ denotes the Legendre transform defined as

$$\mathcal{L}(x, v) := \left(x, \frac{\partial L}{\partial v}(x, v) \right).$$

Under our assumption $\mathcal{L}$ is a diffeomorphism of class at least C^1. We will denote by ϕ_t^L the Euler-Lagrange flow of L, and by X_L the vector field on TM that generates the flow ϕ_t^L.

As done by Mather in [103], it is convenient to introduce for $t > 0$ fixed, the function $h_t : M \times M \to \mathbb{R}$ defined by

$$\forall x, y \in M, \quad h_t(x, y) := \inf \int_0^t L(\gamma(s), \dot{\gamma}(s)) ds,$$

where the infimum is taken over all the absolutely continuous paths $\gamma : [0,t] \to M$ with $\gamma(0) = x$ and $\gamma(t) = y$. The *Peierls barrier* is the function $h : M \times M \to \mathbb{R}$ defined by

$$h(x,y) := \liminf_{t\to\infty} \{h_t(x,y) + c(H)t\} .$$

It is clear that this function satisfies

$$\begin{array}{rcl} \forall x,y,z \in M, \quad h(x,z) & \leq & h(x,y) + h_t(y,z) + c(H)t \\ h(x,z) & \leq & h_t(x,y) + c(H)t + h(y,z). \end{array}$$

Moreover, given a weak KAM solution u, we have

$$u(y) - u(x) \leq h(x,y) \quad \forall x,y \in M.$$

It follows that the function h is either identically $+\infty$ or it is finite everywhere. If M is compact, h is finite everywhere. In addition, if h is finite, then for each $x \in M$, the function $h_x(\cdot) = h(x,\cdot)$ is a critical viscosity solution (see [65] or [69]). Furthermore, h satisfies the triangle inequality

$$\forall x,y,z \in M, \quad h(x,z) \leq h(x,y) + h(y,z).$$

The *projected Aubry set* $\mathcal{A}$ is defined by

$$\mathcal{A} := \{x \in M \mid h(x,x) = 0\} .$$

Since h satisfies the triangle inequality, the function $d_M : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ defined as

$$\forall x,y \in \mathcal{A}, \quad d_M(x,y) := h(x,y) + h(y,x),$$

is a semi-distance on the projected Aubry set. We define the *quotient Aubry set* $(\mathcal{A}_M, d_M)$ to be the metric space obtained by identifying two points in $\mathcal{A}$ if their semi-distance d_M vanishes. In [105], Mather formulated the following problem:

Mather's Problem. If L is C^∞, is the set $\mathcal{A}_M$ totally disconnected, *i.e.* is each connected component of $\mathcal{A}_M$ reduced to a single point?

In [104], Mather brought a positive answer to that problem in low dimension. More precisely, he proved that if M has dimension two, or if the Lagrangian is the kinetic energy associated to a Riemannian metric on M in dimension ≤ 3, then the quotient Aubry set is totally disconnected. In fact, Mather mentioned in [105] that it would be even more interesting to be able to prove that the quotient Aubry set has vanishing one-dimensional Hausdorff measure. The aim of the present chapter is to show that such a property is satisfied under various assumptions. Let us state our results.

Theorem 4.1.1. *If* $\dim M = 1, 2$ *and* H *of class* C^2 *or* $\dim M = 3$ *and* H *of class* $C^{3,1}$, *then* $(\mathcal{A}_M, d_M)$ *has vanishing one-dimensional Hausdorff measure.*

Define the *stationary projected Aubry set* by

$$\mathcal{A}^0 := \{x \in \mathcal{A} \mid (x, d_x h_x) = \mathcal{L}(x, 0)\},$$

and denote by $(\mathcal{A}^0_M, d_M)$ the quotiented metric space. In fact, at the very end of his paper [104], Mather noticed that the argument used in the case where L is a kinetic energy in dimension 3 proves the total disconnectedness of the quotient Aubry set in dimension 3 as long as $\mathcal{A}^0_M$ is empty. Our result concerning the stationary projected Aubry set is the following:

Theorem 4.1.2. *If* $\dim M \geq 3$ *and* H *of class* $C^{k,1}$ *with* $k \geq 2\dim M - 3$, *then* $(\mathcal{A}^0_M, d_M)$ *has vanishing one-dimensional Hausdorff measure. Moreover, if* $\alpha \in (0, 1]$ *is such that* $\alpha(\frac{k+1}{2} + 1) \geq \dim M$ *then* $(\mathcal{A}^0_M, d_M)$ *has vanishing* α*-dimensional Hausdorff measure. In particular, if* H *is* C^∞ *then* $(\mathcal{A}^0_M, d_M)$ *has zero Hausdorff dimension.*

This result is in some sense optimal: for each integer $d > 0$, and each $\epsilon > 0$, Mather has constructed on the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ a Tonelli Lagrangian L of class $C^{2d-3,1-\epsilon}$ such that $\tilde{\mathcal{A}}$ is connected, contained in the fixed points of the Euler-Lagrange flow, and the Mather quotient $(\mathcal{A}_M, d_M)$ is isometric to an interval, see [105].

As a corollary of the above theorem, we have the following result which was moreorless already proved by Mather in [105, §19 page 1722] (see also the work of Sorrentino [124], where the author uses a strategy similar to ours to prove analogous results).

Corollary 4.1.3. *Assume that* H *is of class* C^2 *and that its associated Lagrangian* L *satisfies the following conditions:*

1. $\forall x \in M, \quad \min_{v \in T_x M} L(x, v) = L(x, 0)$;
2. *the mapping* $x \in M \mapsto L(x, 0)$ *is of class* $C^{l,1}(M)$ *with* $l \geq 1$.

If $\dim M = 1, 2$ *or* $\dim M \geq 3$ *and* $l \geq 2\dim M - 3$, *then* $(\mathcal{A}_M, d_M)$ *is totally disconnected. In particular, if* $L(x, v) = \frac{1}{2}\|v\|_x^2 - V(x)$, *with* $V \in C^{l,1}(M)$ *and* $l \geq 2\dim M - 3$ *(*$V \in C^2(M)$ *if* $\dim M = 1, 2$*), then* $(\mathcal{A}_M, d_M)$ *is totally disconnected.*

The Aubry set $\tilde{\mathcal{A}} \subset TM$ can be defined as the set of $(x, v) \in TM$ such that $x \in \mathcal{A}$ and v is the unique $w \in T_x M$ such that $d_x u = \frac{\partial L}{\partial v}(x, w)$ for any critical viscosity subsolution. This set is invariant under the Euler-Lagrange flow ϕ_t^L. Then, for each $x \in \mathcal{A}$, there is only one orbit of ϕ_t^L in $\tilde{\mathcal{A}}$ whose projection passes through x. We denote by $\mathcal{A}^p$ the set of $x \in \mathcal{A}$ whose orbit in the Aubry set is periodic with (strictly) positive period. We call this set the *periodic projected Aubry set*. We have the following result:

Theorem 4.1.4. *If* $\dim M \geq 3$ *and* H *of class* C^k *with* $k \geq 8\dim M - 7$, *then* $(\mathcal{A}_M^p, d_M)$ *has vanishing one-dimensional Hausdorff measure. Moreover, if* $\alpha \in (0,1]$ *is such that* $\alpha(\frac{k-1}{8}+1) \geq \dim M$ *then* $(\mathcal{A}_M^p, d_M)$ *has vanishing* α*-dimensional Hausdorff measure. In particular, if* H *is* C^∞ *then* $(\mathcal{A}_M^p, d_M)$ *has zero Hausdorff dimension.*

In fact, we notice that the method we use to demonstrate Theorem 4.1.4 highlights a general assumption under which we can prove that the quotient Aubry set has small Hausdorff dimension, see Section 4.4. We observe that, if M is assumed to be compact, the size of the quotient Aubry set can be evaluated on the union of the limit sets of the orbits of the Aubry set. Moreover limit sets of flows are well understood on surfaces. Such ideas together with Theorems 4.1.2 and 4.1.4 lead to the following result on surfaces:

Theorem 4.1.5. *If* M *is a compact surface of class* C^∞ *and* H *is of class* C^∞, *then* $(\mathcal{A}_M, d_M)$ *has zero Hausdorff dimension.*

In the last section, we present applications in dynamic of which Theorem 4.1.7 below is a corollary. If X is a C^k vector field on M, with $k \geq 2$, the Mañé Lagrangian $L_X : TM \to \mathbb{R}$ associated to X is defined by

$$L_X(x,v) = \frac{1}{2}\|v - X(x)\|_x^2, \quad \forall (x,v) \in TM,$$

and we denote by ϕ_t^X the flow generated by X. We now recall the definition of chain-recurrence:

Definition 4.1.6. Take $\varepsilon > 0$. A *periodic* ε*-chain* is a sequence $(x_0, \dots, x_n)$ such that $x_0 = x_n$ and there exist $t_i \geq 1$ such that $d(x_{i+1}, \phi_{t_i}^X(x_i)) \leq \varepsilon$ for $i = 0, \dots, n-1$. A point x is *chain-recurrent* if for any $\varepsilon > 0$ there exists a periodic ε-chain passing through x.

Roughly speaking, the chain-recurrence and the fact of being in the projected Aubry set are two different ways to characterize the set of points which are more or less invariant under the flow for large times. It is therefore a natural question to understand when these two definitions coincide. Fathi has raised the following problem (compare with the list of questions *http://www.aimath.org/WWN/dynpde/articles/html/20a/*):

Problem. Let $L_X : TM \to \mathbb{R}$ be the Mañé Lagrangian associated to the $C^k, k \geq 2$, vector field X on the compact connected manifold M.

(1) Is the set of chain-recurrent points of the flow of X on M equal to the projected Aubry set $\mathcal{A}$?

(2) Give a condition on the dynamics of X that ensures that the only weak KAM solutions are the constants.

The theorems obtained in the first part of the chapter together with applications in dynamics developed in Section 6 give an answer to this question if $\dim M \leq 3$.

Theorem 4.1.7. *Let X be a $\mathrm{C}^k, k \geq 2$ vector field on the compact connected C^∞ manifold M. Assume that one of the conditions hold:*

(1) The dimension of M is 1 or 2.

(2) The dimension of M is 3, and the vector field X never vanishes.

(3) The dimension of M is 3, and X is of class $\mathrm{C}^{3,1}$.

Then the projected Aubry set $\mathcal{A}$ of the Mañé Lagrangian $L_X : TM \to \mathbb{R}$ associated to X is the set of chain-recurrent points of the flow of X on M. Moreover, the constants are the only weak KAM solutions for L_X if and only if every point of M is chain-recurrent under the flow of X.

The outline is the following: Sections 2 and 3 are devoted to preparatory results. Section 4 is devoted to the proofs of Theorems 4.1.1, 4.1.2 and 4.1.4. Sections 5 and 6 present applications in dynamics.

4.2 Preparatory lemmas

We denote by $\mathcal{SS}$ the set of critical viscosity subsolutions and by $\mathcal{S}_-$ the set of weak KAM solutions. Hence $\mathcal{S}_- \subset \mathcal{SS}$. If $u : M \to \mathbb{R}$ is a critical viscosity subsolution, we recall that

$$u(y) - u(x) \leq h(x, y), \quad \forall x, y \in M.$$

In [69], Fathi and Siconolfi proved that for every critical viscosity subsolution $u : M \to \mathbb{R}$, there exists a C^1 critical subsolution whose restriction to the projected Aubry set is equal to u. In the sequel, we denote by $\mathcal{SS}^1$ the set of C^1 critical subsolutions. The following lemma is fundamental in the proof of our results.

Lemma 4.2.1. *For every $x, y \in \mathcal{A}$,*

$$\begin{aligned} d_M(x,y) &= \max_{u_1,u_2 \in \mathcal{S}_-} \{(u_1 - u_2)(y) - (u_1 - u_2)(x)\} \\ &= \max_{u_1,u_2 \in \mathcal{SS}} \{(u_1 - u_2)(y) - (u_1 - u_2)(x)\} \\ &= \max_{u_1,u_2 \in \mathcal{SS}^1} \{(u_1 - u_2)(y) - (u_1 - u_2)(x)\}. \end{aligned}$$

Proof. Let $x, y \in \mathcal{A}$ be fixed. First, we notice that if u_1, u_2 are two critical viscosity subsolutions, then we have

$$\begin{aligned}(u_1 - u_2)(y) - (u_1 - u_2)(x) &= (u_1(y) - u_1(x)) + (u_2(x) - u_2(y)) \\ &\leq h(x, y) + h(y, x) = d_M(x, y).\end{aligned}$$

Moreover, if we define $u_1, u_2 : M \to \mathbb{R}$ by $u_1(z) := h(x, z)$ and $u_2(z) := h(y, z)$ for any $z \in M$, then we have

$$\begin{aligned}(u_1 - u_2)(y) - (u_1 - u_2)(x) &= (h(x, y) - h(y, y)) - (h(x, x) - h(y, x)) \\ &= h(x, y) + h(y, x) = d_M(x, y),\end{aligned}$$

since $h(x, x) = h(y, y) = 0$. Since u_1, u_2 are both critical viscosity solutions, we obtain easily the first and the second equality. The last inequality is an immediate consequence of the Theorem of Fathi and Siconolfi recalled above. □

The proofs of the next two Lemmas can be found in [114]. It has to be noticed that Norton's Lemma 4.2.2 is an elegant generalization of the Morse original Lemma, see [111].

Lemma 4.2.2 (The generalized Morse Vanishing Lemma). *Let $k \in \mathbb{N}$ and $\alpha \in [0, 1]$. Then any set $A \subset \mathbb{R}^n$ can be decomposed into a countable union $A = \cup_{i \in \mathbb{N}} A_i$ where*

1. *A_0 is countable;*
2. *$A_i \subset B_i$ with B_i a C^1-embedded compact disk of dimension $\leq n$ such that every $f \in C^{k,\alpha}(\mathbb{R}^n, \mathbb{R})$ vanishing on A satisfies, for each $i \geq 1$,*

$$|f(x) - f(y)| \leq M_i |x - y|^{k+\alpha} \quad \forall y \in A_i,\ x \in B_i \tag{4.2.1}$$

for a certain constant M_i.

Lemma 4.2.3. *For any C^1-embedded compact disk B, there is a constant $C > 0$ such that for all $x, y \in B$ there is a C^1 path in B from x to y with length less than $C|x - y|$.*

The proof of Lemma 4.2.4 that we present here is derived from [18] (compare [72]) who proved that if $E \subset \mathbb{R}^n$ is a measurable set, $f : E \to \mathbb{R}$ is continuous, and $n \geq 2$ is such that f satisfies

$$|f(x) - f(y)| \leq C|x - y|^n \quad \forall x, y \in E,$$

then $f(E)$ has Lebesgue measure zero.

Lemma 4.2.4. *Let $\Psi : E \to X$ be a map where E is a subset of $\mathbb{R}^n$ and (X, d_X) is a semi-metric space. Suppose that there are $\alpha \in (0,1]$ and $M > 0$ such that*

$$\forall x, y \in E, \quad d_X(\Psi(x), \Psi(y)) \leq M|x-y|^{\frac{n}{\alpha}}, \tag{4.2.2}$$

then the α-dimensional Hausdorff measure of $(\Psi(E), d_X)$ is zero.

Proof. Since it suffices to prove that $\mathscr{H}^\alpha(\Psi(E \cap K)) = 0$ for each compact set $K \subset \mathbb{R}^n$, we can assume that E is bounded, which in particular implies $\mathscr{L}^n(E) < +\infty$. We now write $E = E_1 \cup E_2$, where E_1 is the set of density points for E and $E_2 := E \setminus E_1$. It is a standard result in measure theory that $\mathscr{L}^n(E_2) = 0$. Thus for each $\epsilon > 0$ be fixed, there exists a countable family of balls $\{B_i\}_{i \in I}$ such that

$$E_2 \subset \bigcup_{i \in I} B_i \quad \text{and} \quad \sum_{i \in I} (\operatorname{diam} B_i)^n \leq \varepsilon.$$

Then we have

$$\mathscr{H}^\alpha(\Psi(E_2)) \leq \sum_{i \in I} (\operatorname{diam}_X \Psi(B_i))^\alpha \leq M \sum_{i \in I} (\operatorname{diam} B_i)^n \leq M\varepsilon.$$

Letting $\varepsilon \to 0$, we obtain $\mathscr{H}^\alpha(\Psi(E_2)) = 0$.
We now want to prove that $\mathscr{H}^\alpha(\Psi(E_1)) = 0$. Fix $P \in \mathbb{N}$. For every density point $x \in E_1$, there exists $\rho = \rho(x) > 0$ such that

$$\frac{\mathscr{L}^n(E_1 \cap B(x,r))}{\mathscr{L}^n(B(x,r))} = \frac{\mathscr{L}^n(E \cap B(x,r))}{\mathscr{L}^n(B(x,r))} \geq 1 - \frac{1}{2P^n} \quad \forall r \leq \rho(x). \tag{4.2.3}$$

Now it is simple to prove that for all $y, z \in E_1 \cap B(x,r)$ there exist $P+1$ points $x_0, \dots, x_P \in E_1$, with $x_0 = y$ and $x_P = z$ such that

$$|x_i - x_{i-1}| \leq \frac{4r}{P} \quad \forall 1 \leq i \leq P$$

Indeed, first take $y_1, \dots, y_{P-1}$ the $P-1$ points on the line segment $[y, x]$ such that $|y_i - y_{i-1}| = \frac{|y-x|}{P}$ and then observe that, by (4.2.3), $B(y_i, \frac{r_x}{P}) \cap E_1$ is not empty for each i, and so it suffices to take a point x_i in that set. Then,

$$\begin{aligned} d_X(\Psi(y), \Psi(z)) &\leq \sum_{i=1}^P d_X(\Psi(x_{i-1}), \Psi(x_i)) \leq M \sum_{i=1}^P |x_i - x_{i-1}|^{\frac{n}{\alpha}} \\ &\leq MP\left(\frac{4r}{P}\right)^{\frac{n}{\alpha}} = 4^{\frac{n}{\alpha}} M P^{1-\frac{n}{\alpha}} r^{\frac{n}{\alpha}}. \end{aligned} \tag{4.2.4}$$

We are now able to prove that $\mathscr{H}^\alpha(\Psi(E_1)) = 0$.
Take an open set $\Omega \supset E_1$ such that $\mathscr{L}^n(\Omega) \le 2\mathscr{L}^n(E_1) = 2\mathscr{L}^n(E) < +\infty$ (we can assume $\mathscr{L}^n(E) > 0$) and consider the fine covering $\mathcal{F}$ given by $\mathcal{F} = \{B(x,r)\}_{x\in F_1}$ with r such that $B(x,r) \subset \Omega$ and $r \le \frac{\rho(x)}{5}$, where $\rho(x)$ was defined above. By Vitali's covering theorem (see [61, paragraph 1.5.1]), there exists a countable collection $\mathscr{G}$ of disjoint balls in $\mathcal{F}$ such that

$$E_2 \subset \bigcup_{B\in\mathscr{G}} 5B,$$

where $5B$ denotes the ball concentric to B with radius 5 times that of B. We can so consider the covering of $f(E_1)$ given by $\cup_{1\le i\le N_n}\{f(5B\cap E_1)\}_{B\in\mathscr{G}}$. In this way, by (4.2.4), we get

$$\begin{aligned}\mathscr{H}^\alpha(\Psi(E_1)) &\le \sum_{B\in\mathscr{G}} (\operatorname{diam}_X \Psi(5B\cap E_1))^\alpha \le 4^n M^\alpha P^{\alpha-n} \sum_{B\in\mathscr{G}} (\operatorname{diam} 5B)^n \\ &\le 20^n M^\alpha P^{\alpha-n} \sum_{B\in\mathscr{G}} (\operatorname{diam} B)^n \\ &\le 20^n M^\alpha P^{\alpha-n} \mathscr{L}^n(\Omega) \le 2\cdot 20^n M^\alpha P^{\alpha-n}\mathscr{L}^n(E),\end{aligned}$$

and, in this case we conclude letting $P \to +\infty$, as $n \ge 2$. □

4.3 Existence of $C^{1,1}_{loc}$ critical subsolution on noncompact manifolds

In [20], using some kind of Lasry-Lions regularization (see [89]), Bernard proved the existence of $C^{1,1}$ critical subsolutions on compact manifolds. Here, adapting his proof, we show that the same result holds in the noncompact case and we make clear that the Lipschitz constant of the derivative of the $C^{1,1}_{loc}$ critical subsolution can be uniformly bounded on compact subsets of M.

Theorem 4.3.1. *Assume that H is of class C^2. For every open subset $\mathcal{O}$ of M which is relatively compact in M, there is a constant $L = L(\mathcal{O}) > 0$ such that if $u : M \to \mathbb{R}$ is a critical viscosity subsolution, then there exists a $C^{1,1}_{loc}$ critical subsolution $v : M \to \mathbb{R}$ whose restriction to the projected Aubry set is equal to u and such that the mapping $x \mapsto d_x v$ is L-Lipschitz on $\mathcal{O}$.*

Before proving Theorem 4.3.1, we observe that the following result holds:

Lemma 4.3.2. *There is a constant $K' > 0$ such that any critical viscosity subsolution $u : M \to \mathbb{R}$ is K'-Lipschitz on M, that is,*

$$|u(y) - u(x)| \leq K'd(x, y), \quad \forall x, y \in M,$$

where d denotes the Riemannian distance associated to the metric g.

Proof. Let $u : M \to \mathbb{R}$ be a critical viscosity subsolution and $x, y \in M$ be fixed. Let $\gamma_{x,y} : [0, d(x, y)] \to M$ be a minimizing geodesic with constant unit speed joining x to y. By definition of $h_{d(x,y)}(x, y)$, one has

$$h_{d(x,y)}(x, y) \leq \int_0^{d(x,y)} L(\gamma_{x,y}(t), \dot{\gamma}_{x,y}(t))\, dt \leq A(1)d(x, y),$$

where $A(1) := \sup_{x\in M}\{L(x, v) \mid \|v\|_x \leq 1\}$ is finite thanks to the uniform superlinearity of H in the fibers. Thus, one has

$$u(x) - u(y) \leq h_{d(x,y)}(x, y) + c(H)d(x, y) \leq (A(1) + c(H))d(x, y).$$

Exchanging x and y, we conclude that u is K'-Lipschitz with $K' := A(1) + c(H)$. □

Before giving the proof of the theorem, we also notice that since L is uniformly superlinear in the fibers, there exists a finite constant $C(K')$ such that

$$L(x, v) \geq 2K'\|v\|_x + C(K') \quad \forall (x, v) \in TM.$$

From that, we deduce that for every $t > 0$,

$$h_t(x, y), h_t(y, x) \geq 2K'd(x, y) + C(K')t, \quad \forall x, y \in M. \tag{4.3.1}$$

Proof of Theorem 4.3.1. Let K_n be an increasing sequence of compact sets such that $K_n \subset \overset{\circ}{K}_{n+1}$ and $\cup_n K_n = M$. We consider the two Lax-Oleinik semigroups T_t^- and T_t^+ defined by

$$T_t^- u(x) := \inf_{y\in M} \{u(y) + h_t(y, x) + c(H)t\}, \quad T_t^+ u(x) := \sup_{y\in M} \{u(y) - h_t(x, y) - c(H)t\},$$

for every $x \in M$. In [65], Fathi proved that those two semigroups preserve the set of critical viscosity subsolutions and that, for all $t > 0$ and each continuous function u, the function $T_t^+ u$ is locally semi-convex, while $T_t^- u$ is locally semi-concave. In [20], the idea for proving the existence of $C^{1,1}$ critical subsolution on compact manifolds is the following. It is a known fact that a function is $C^{1,1}$ if and only if it is both semi-concave and semi-convex. Let now u be a critical viscosity subsolution. If we apply the

semigroup T_t^+ to u, we obtain a semi-convex critical viscosity subsolution T_t^+u. Thus, if one proves that, for s small enough, $T_s^-T_t^+u$ is still semi-convex, as we already know that it is semi-concave, we would have found a $C^{1,1}$ critical subsolution. Since we want to give a uniform bound on the Lipschitz constant of the derivative of the $C^{1,1}_{loc}$ critical subsolution on compact sets, we will have to bound the constant of semi-convexity of T_t^+u on compact subsets of M. Let us now prove the result in the noncompact case.
Let $u : M \to \mathbb{R}$ be a critical viscosity subsolution. Let $t > 0$ be fixed, we set $v := T_t^+u$.

Lemma 4.3.3. *The function $v : M \to \mathbb{R}$ is locally semi-convex on M. Morever, there exists a bounded subset F of $C^2\left(\overset{\circ}{K}_4, \mathbb{R}\right)$ whose bound depends only on t (not on u) verifying*

$$v(x) = \max_{f\in F} f(x), \quad \forall x \in K_4,$$

and such that for each $x \in K_4$ and each $p \in D^-v(x)$ there is $f \in F$ such that $f(x) = v(x)$ and $d_xf = p$.

Proof. Let $x \in K_4$ be fixed. From the definition of $T_t^+u(x)$, we have

$$\begin{aligned} T_t^+u(x) &\geq u(x) - h_t(x,x) - c(H)t \\ &\geq u(x) - tL(x,0) - c(H)t \\ &\geq u(x) - (A(0) + c(H))t, \end{aligned}$$

where $A(0) := \sup_{x\in M}\{L(x,0)\}$ is finite thanks to the uniform superlinearity of H in the fibers. On the other hand, by Lemma 4.3.2 and (4.3.1), we have for every $y \in M$,

$$\begin{aligned} u(y) - h_t(x,y) - c(H)t &\leq u(x) + K'd(x,y) - 2K'd(x,y) - c(K')t - c(H)t \\ &\leq u(x) - K'd(x,y) - (C(K') + c(H))t. \end{aligned}$$

Therefore, the supremum in the definition of $T_t^+u(x)$ is necessarily attained at a point $y_x \in M$ satisfying

$$d(x, y_x) \leq \frac{(A(0) - C(K'))}{K'}t.$$

Denote by K_x the set of $y \in M$ such that $d(x,y) \leq \frac{(A(0)-C(K'))}{K'}t$, and by K the union of K_x for $x \in K_4$. K is a compact subset of M. By Proposition 6.2.17 of the Appendix, there is a compact set $\tilde{K} \subset M$ and a constant $A > 0$ such that every curve $\gamma : [0,t] \to M$ with $\gamma(0), \gamma(t) \in K$ and $\int_0^t L(\gamma(s), \dot\gamma(s))\, ds = h_t(\gamma(0), \gamma(t))$ satisfies

$$\gamma\left([0,t]\right) \subset \tilde{K}, \quad \text{and} \quad \|\dot\gamma(s)\|_{\gamma(s)} \leq A \quad \forall s \in [0,t]. \tag{4.3.2}$$

Let $x \in K_4$. By construction of K_x, there is $y_x \in K_x$ such that

$$T_t^+u(x) = u(y_x) - \int_0^t L(\gamma_x(s), \dot\gamma_x(s))ds - c(H)t,$$

where $\gamma_x : [0,t] \to M$ is a curve such that $\gamma_x(0) = x, \gamma_x(t) = y_x$ and

$$h_t(x, y_x) = \int_0^t L(\gamma_x(s), \dot\gamma_x(s))ds$$

Now, for any $x \in K_4$ and $y \in K_x$, there are $\mathcal{V}_x$ an open neighbourhood of x and $\Gamma_{x,y} : \mathcal{V}_x \times [0,t] \to M$ a smooth mapping such that $\Gamma_{x,y_x}(x,\cdot) = \gamma_x(\cdot)$, and such that for every $x' \in \mathcal{V}_x$ and $y \in K_x$, $\Gamma_{x,y}(x',\cdot)$ is a smooth curve joining x' to y. We have

$$\begin{aligned} T_t^+ u(x') &\geq u(y_x) - \int_0^t L\left(\Gamma_{x,y_x}(x',s), \frac{d\Gamma_{x,y_x}}{ds}(x',s)\right) ds - c(H)t \\ &\geq T_t^+ u(x) + \phi_{x,y_x}(x'), \end{aligned}$$

where the function $\phi_x : \mathcal{V}_x \to \mathbb{R}$ is defined by

$$\phi_{x,y_x}(x') := \int_0^t L\left(\Gamma_{x,y_x}(x,s), \frac{d\Gamma_{x,y_x}}{ds}(x,s)\right) - L\left(\Gamma_{x,y_x}(x',s), \frac{d\Gamma_{x,y_x}}{ds}(x',s)\right) ds$$

for all $x' \in \mathcal{V}_x$. The function ϕ_{x,y_x} is smooth and satisfies $\phi_{x,y_x}(x) = 0$. By construction (because of the compactness of the set $\tilde{K}$), it is clear that the set of functions $\{\phi_{x,y_x}\}_{x\in K_4}$ can be bounded in $C^2\left(\overset{\circ}{K}_4\right)$. More in general, the whole family $G := \{\phi_{x,y}\}_{x\in K_4, y\in K_x}$ is bounded in C^2.

Since K_4 is compact, up to working in local charts, and using standard arguments to extend the C^2 functions of our family constructed in charts to an open neighborhood of K_4 in such a way to preserve a global C^2 bound, we can assume that we are in $\mathbb{R}^n$. Thus, by [119, Proposition 6] applied to $-T_t^+ u$, we obtain that $v = T_t^+ u$ is σ-semiconvex on K_4, with the constant σ depending only on the C^2 bound of the family G (and therefore is independent of the subsolution u). Now, by [119, Proposition 7], for any $x \in K_4$ and any $p \in D^- v(x)$ there exists a parabola $P_{x,p}$ with second derivative bounded by σ which touches v from below at x with $d_x P_{x,p} = p$. By Lemma 4.3.2 we have the global bound $\|p\|_x \leq K'$, and therefore $F := \{P_{x,p}\}$ is the desired family. □

We claim that, for $t_1, s_1 > 0$ small enough, the function $u_1 := T_{s_1}^- T_{t_1}^+ u$ is $C^{1,1}$ on K_2 and that the Lipschitz constant of its derivative can be bounded independently of u. In order to prove this claim, we will show that, for s small enough, we have

$$T_s^- T_t^+ u(x) := \min_{y\in K_3} \left\{T_t^+ u(y) + h_s(y,x) + c(H)s\right\}, \quad \forall x \in K_2. \tag{4.3.3}$$

Once we will have proved this, the problem of proving $C^{1,1}$ regularity in K_2 will be exactly the same as in the compact case and so the proof in [20] will work.

Indeed, always as in [20], for s small enough $T_s^-(F)$ is a bounded subset of $C^2\left(\mathring{K}_3\right)$ and, by (4.3.3), one can write

$$T_s^- v = \max_{f\in F} T_s^- f \quad \text{on } K_2,$$

that implies that $T_s^- T_t^+ u$ is $C^{1,1}$ on K_2. Moreover, we can assume that s is sufficiently small so that $T_s^-(F)$ is bounded in C^2 by a constant σ' which is still independent of u, and this implies that the $C^{1,1}$ bound is independent of the particular subsolution u. Let us now prove (4.3.3).

Set $v := T_t^+ u$ and fix $s > 0$. We recall that v is critical viscosity subsolution. Since v is K'-Lipschitz on M, we deduce that for any $x, y \in M$,

$$\begin{aligned} v(y) + h_s(y,x) + c(H)s &\geq v(x) - K'd(x,y) + 2K'd(x,y) + C(K')s + c(H)s \\ &\geq v(x) + K'd(x,y) + (C(K') + C(H))s. \end{aligned}$$

But, taking $y = x$ in the formula defining $T_s^- v(x)$ yields for any $x \in M$,

$$\begin{aligned} T_s^- v(x) &\leq v(x) + h_s(x,x) + c(H)s \\ &\leq v(x) + sL(x,0) + c(H)s \leq v(x) + (A(0) + c(H))s. \end{aligned}$$

Consequently, we deduce that, for every $x \in M$, the infimum in the definition of $T_s^- v(x)$ is necessarily attained at a point $y_x \in M$ satisfying

$$d(x, y_x) \leq \frac{(A(0) - C(K'))}{K'} s.$$

So (4.3.3) follows taking s such that $\frac{(C-c(H))}{A(0)+c(H)} s \leq \mathrm{dist}(K_2, K_3^c)$. As we said above, now the proof given in [20] allows us to say that u_1 is $C^{1,1}$ in K_2. Let us now define $u_2(x) := T_{s_2}^- T_{t_2}^+ u_1(x)$. Arguing as above we get that, for t_2, s_2 smalls enough, u_2 is $C^{1,1}$ in K_3. We now claim that, taking t_2, s_2 sufficiently smalls, we also have that

$$Lip_{K_1}(d_x u_2) \leq \left(1 + \frac{1}{2}\right) Lip_{K_2}(d_x u_1), \tag{4.3.4}$$

where, for a function f, we denote by $Lip_{K_n}(f)$ the Lipschitz constant of f on K_n. This simply follows observing that, if we denote by $\Gamma_{u_2} \subset T^*M$ the graph of the differential of u_2, as $d_x u_2$ is Lipschitz on K_2, the evolution of u_2 in K_1 by the two semigroups corresponds to the evolution of Γ_{u_2} in T^*K_1 by the Hamiltonian flow, at least for small times (see [20]). Thus the smoothness of the Hamiltonian flow tells us that the Lipschitz constant of $d_x u_2$ uniformly converges in K_1 to the Lipschitz constant of $d_x u_1$. In particular, for t_2, s_2 sufficiently smalls, we get (4.3.4). Now we iterate the construction in the following way:

$$u_{n+1}(x) := T^-_{s_{n+1}} T^+_{t_{n+1}} u_n(x),$$

with t_{n+1}, s_{n+1} smalls enough so that

$$u_{n+1} \text{ is } C^{1,1} \text{ in } K_{n+2},$$

$$Lip_{K_n}(d_x u_{n+1}) \leq \left(1 + \frac{1}{2^n}\right) Lip_{K_{n+1}}(d_x u_n).$$

In this way, by Ascoli-Arzelà theorem and a diagonal argument, we find a subsequence u_{n_k} which converges in the C^1 topology to a function $u_\infty \in C^1$. Moreover, as

$$\prod_{n=1}^{\infty} \left(1 + \frac{1}{2^n}\right) < +\infty,$$

for a fixed compact K_l we have an uniform bound for the Lipschitz constant of $d_x u_n$ on K_l for any n. This implies that $u_\infty \in C^{1,1}_{loc}$. □

4.4 Proofs of Theorems 4.1.1, 4.1.2, 4.1.4

4.4.1 Proof of Theorem 4.1.1

Let us first assume that $\dim M = 1, 2$. Without loss of generality, we can assume that we work in a relatively compact open set $\mathcal{O} \subset \mathbb{R}^n$ with $n = 1, 2$. Our aim is to apply Lemma 4.2.4 with $E = \mathcal{A} \cap \mathcal{O}$, $(X, d_X) = (M, d_M)$, $\Psi = Id$ and $\alpha = 1$. By Lemma 4.2.1, we know that for every $x, y \in \mathcal{A}$,

$$d_M(x, y) = \max_{u_1, u_2 \in \mathcal{S}_-} \left\{ (u_1 - u_2)(y) - (u_1 - u_2)(x) \right\}.$$

Let $u_1, u_2 : M \to \mathbb{R}$ be two weak KAM solutions be fixed. It is well known that both the mappings $x \in M \mapsto d_x u_1$, $x \in M \mapsto d_x u_2$ coincide and are locally Lipschitz on $\mathcal{A}$ (see [65]). Thus there is a constant $C > 0$ which does not depend on u_1 and u_2 such that if we set $v := u_1 - u_2$, we have

$$|v(y) - v(x)| \leq C|x - y|^2, \quad \forall x, y \in \mathcal{A} \cap \mathcal{O}.$$

Since u_1 and u_2 are arbitrary, we get

$$d_M(x, y) \leq C|x - y|^2, \quad \forall x, y \in \mathcal{A} \cap \mathcal{O}.$$

By Lemma 4.2.4, we deduce that if $\dim M = 1, 2$, $(\mathcal{A}_M, d_M)$ has vanishing one-dimensional Hausdorff measure.

Let us now assume that $\dim M = 3$. The fact that $\mathcal{A}_M^0$ has vanishing one-dimensional Hausdorff measure will follow from Theorem 4.1.2. So, it suffices to prove that the semi-metric space $\mathcal{A} \setminus \mathcal{A}^0$ has vanishing one-dimensional Hausdorff measure. Set $\mathcal{A}' := \mathcal{A} \setminus \mathcal{A}^0$ and consider $x \in \mathcal{A}'$. From [69, Proposition 5.2], there exists a curve $\gamma : \mathbb{R} \to M$ such that $\gamma(0) = x$ and such that each critical viscosity subsolution $u : M \to \mathbb{R}$ satisfies

$$u(\gamma(t')) - u(\gamma(t)) = \int_t^{t'} L(\gamma(s), \dot\gamma(s))ds + c(H)(t' - t),$$

for all $t < t' \in \mathbb{R}$. In particular, each such u is differentiable at each point of γ and satisfies

$$d_{\gamma(t)}u = \frac{\partial L}{\partial v}(\gamma(t), \dot\gamma(t)), \quad \forall t \in \mathbb{R}$$

(see [65]). Therefore we deduce that the function $v : M \to \mathbb{R}$ defined as $v := u_1 - u_2$ is constant along the orbits of the Aubry set. Since this is true for any pair of KAM solutions, this implies that $d_M(\gamma(s), \gamma(t)) = 0$ for all $s, t \in \mathbb{R}$. As a consequence, it suffices to prove that the set $(\mathcal{A}' \cap S, d_M)$ has vanishing one-dimensional Hausdorff measure for each open surface $S \subset M$ which is locally transverse to the orbits of the Aubry set. As above, this is a consequence of the fact that the mapping $x \mapsto d_x u$ is locally Lipschitz on $\mathcal{A}$ (with a constant which does not depend on u) and Lemma 4.2.4.

4.4.2 Proof of Theorem 4.1.2

Without loss of generality, we can assume that we work on a relatively compact open set $\mathcal{O} \subset \mathbb{R}^n$ and that $c(H) = 0$. Set for every $x \in \mathcal{O}$,

$$\tilde H(x) := \min_{p \in \mathbb{R}^n} \{H(x, p)\} = -L(x, 0),$$

and

$$\tilde p(x) := \frac{\partial L}{\partial v}(x, 0).$$

The function $\tilde H$ is of class $C^{k,1}$ on $\mathcal{O}$ and satisfies for every $x \in \mathcal{O}$,

$$\tilde H(x) = H(x, \tilde p(x)) \le 0.$$

Moreover, we notice that, by strict convexity of H and the fact that $\mathcal{O}$ is relatively compact, there exists $\alpha \ge 0$ such that for every $x \in \mathcal{O}$,

$$H(x, p) \le 0 \Longrightarrow |p - \tilde p(x)| \le \alpha \sqrt{-\tilde H(x)}. \tag{4.4.1}$$

Using lemma 4.2.2, we can decompose $\mathcal{A}^0$ as

$$\mathcal{A}^0 = \cup_{i\in\mathbb{N}} A_i.$$

First, since A_0 is countable, it has zero Hausdorff dimension. Let us now show that each A_i has vanishing one-dimensional Hausdorff measure. Let $i \geq 1$ be fixed. Since $\tilde{H}$ is a $C^{k,1}$ function vanishing on $\mathcal{A}^0$, by (4.2.1) we know that

$$-\tilde{H}(x) = \left|\tilde{H}(x) - \tilde{H}(y)\right| \leq M_i|x-y|^{k+1}, \quad \forall y \in A_i, \quad \forall x \in B_i.$$

Hence, from (4.4.1), we have for every C^1 critical subsolution $u : M \to \mathbb{R}$,

$$|d_x u - \tilde{p}(x)| \leq \alpha\sqrt{M_i}|x-y|^{\frac{k+1}{2}}, \quad \forall y \in A_i, \quad \forall x \in B_i,$$

which gives for every pair of C^1 critical subsolutions $u_1, u_2 : M \to \mathbb{R}$,

$$|d_x(u_1-u_2)| \leq C_i|x-y|^{\frac{k+1}{2}}, \quad \forall y \in A_i, \quad \forall x \in B_i,$$

for a certain constant C_i. By lemma 4.2.3, integrating along a path from x to y in B_i yields

$$|(u_1-u_2)(y) - (u_1-u_2)(x)| \leq \tilde{C}C_i|x-y|^{\frac{k+1}{2}+1}, \quad \forall x, y \in A_i.$$

By Lemma 4.2.1, we deduce that

$$d_M(x,y) \leq \tilde{C}C_i|x-y|^{\frac{k+1}{2}+1}, \quad \forall x, y \in A_i.$$

In order to conclude that A_i has vanishing one-dimensional Hausdorff measure, it suffices, from Lemma 4.2.4, to have k such that $\frac{k+1}{2}+1 \geq n$, that is $k \geq 2n-3$. As a consequence, we proved that the one-dimensional Hausdorff dimension of $(\mathcal{A}^0_M, d_M)$ vanishes as soon as $k \geq 2n-3$. Finally, the fact that the α-dimensional Hausdorff measure of $(\mathcal{A}^0_M, d_M)$ vanishes whenever $\alpha \in (0,1]$ is such that $\alpha(\frac{k+1}{2}+1) \geq \dim M$ follows from the same arguments.

4.4.3 Proof of Theorem 4.1.4

Let $\tilde{\mathcal{A}}^p$ be the set of points in the Aubry set $\tilde{\mathcal{A}}$ that project on $\mathcal{A}^p$. For every $x \in \mathcal{A}$, let us denote by $T(x)$ the period of the unique orbit of the Euler-Lagrange flow in $\tilde{A}$ passing through x. Fix $\bar{x} \in \mathcal{A}^p$ and denote by $\bar{\gamma} : [0, T(\bar{x})] \to M$ the path such that $\bar{\gamma}(0) = \bar{\gamma}(T(\bar{x})) = \bar{x}$ and satisfying

$$h_{T(\bar{x})}(\bar{x},\bar{x}) + c(H)T(\bar{x}) = \int_0^{T(\bar{x})} L(\bar{\gamma}(s),\dot{\bar{\gamma}}(s))ds + c(H)T(\bar{x}) = 0.$$

Let S be a smooth hypersurface in M which is locally transverse to $\bar{\gamma}$ at $\bar{x}$ and E be the fiber bundle over S, that is, the set of $(x, v) \in TM$ with $x \in S$. For every $(x, v) \in E$, let $\tau(x, v)$ be the first time $t > 0$ such that $\phi_t(x, v) \in E$. For sake of simplicity, for every $(x, v) \in E$, we denote by $\gamma_{x,v} : [0, \tau(x, v)] \to M$ the trajectory of the Euler-Lagrange flow starting at (x, v). We define the mapping $\theta : E \to S$ by

$$\theta(x, v) := \gamma_{x,v}(\tau(x, v)), \quad \forall (x, v) \in E.$$

So θ is something like the Poincaré map (or first return map) associated with $\bar{\gamma}$ and S, it is well defined in a small neighbourhood $N \subset E$ of $(\bar{x}, \bar{v})$ and it is of class C^{k-1}. Denote by $d_g^S(\cdot, \cdot)$ the distance on S which corresponds to the Riemannian metric induced by g on S. We recall that the mapping $(x, y) \mapsto d_g^S(x, y)^2$ is smooth in a small neighbourhood N_S of $\bar{x}$ in S. Without loss of generality we can assume that $\theta(N) \subset N_S$. Define the mapping $\Psi : N \to \mathbb{R}$ by

$$\Psi(x, v) := \Big(\int_0^{\tau(x,v)} L(\gamma_{x,v}(s), \dot{\gamma}_{x,v}(s)) ds + c(H)\tau(x, v) \Big)^2 + d_g^S(\theta(x, v), x)^2, \quad \forall (x, v) \in N.$$

By construction, there exists $\delta(\bar{x}) > 0$ such that, for every $(x, v) \in \tilde{\mathcal{A}}^p \cap N$, we have

$$T(x) \in (\bar{T} - \delta(\bar{x}), \bar{T} + \delta(\bar{x})) \Longrightarrow \Psi(x, v) = 0,$$

where $\bar{T} := T(\bar{x})$. Denote by $\tilde{A}_{\bar{x}}$ the set of $(x, v) \in \tilde{\mathcal{A}}^p \cap N$ such that $T(x) \in (\bar{T} - \delta(\bar{x}), \bar{T} + \delta(\bar{x}))$, and by $A_{\bar{x}}$ its projection on M. Furthermore, we notice that for every $(x, v) \in N$, if we consider a minimizing geodesic with unit speed (for the Riemannian metric g on S) $\gamma : [0, d_g^S(\theta(x, v), x)] \to S$ joining $\theta(x, v)$ to x, we have

$$h_{\rho(x,v)}(x, x) \leq \int_0^{\tau(x,v)} L(\gamma_{x,v}(s), \dot{\gamma}_{x,v}(s)) ds + \int_0^{d_g^S(\theta(x,v),x)} L(\gamma(s), \dot{\gamma}(s)) ds,$$

where $\rho(x, v)$ is defined by

$$\rho(x, v) := \tau(x, v) + d_g^S(\theta(x, v), x), \quad \forall (x, v) \in N.$$

Hence, if we denote by J an upper bound for $|L(x, v)|$ with $x \in N_S$ and $v \in T_x S$ satisfying $|v|_g \leq 1$, we obtain for every $(x, v) \in N$,

$$\begin{aligned} h_{\rho(x,v)}(x, x) + c(H)\rho(x, v) \;&\leq\; \Big| \int_0^{\tau(x,v)} L(\gamma_{x,v}(s), \dot{\gamma}_{x,v}(s)) ds \\ &+\; c(H)\tau(x, v) \Big| + (c(H) + J) d_g^S(\theta(x, v), x) \\ &\leq\; (1 + c(H) + J)\sqrt{\Psi(x, v)}. \end{aligned} \tag{4.4.2}$$

Without loss of generality, we can assume from now that we work in $\mathbb{R}^n$. From Lemma 4.2.2, we can decompose the set $\tilde{A}_{\bar{x}}$ as

$$\tilde{A}_{\bar{x}} = \cup_{i \in \mathbb{N}} \tilde{A}_i.$$

Let $i \geq 1$ be fixed. Since Ψ is of class C^{k-1} on N, by (4.2.1) we know that

$$0 \leq \Psi(x,v) \leq M_i \left(|x-y|^2 + |v-w|^2\right)^{\frac{k-1}{2}}, \quad \forall (y,w) \in \tilde{A}_i, \quad \forall (x,v) \in \tilde{B}_i. \quad (4.4.3)$$

We need now the following result.

Lemma 4.4.1. *Let $L > 0$, $t_0 > 0$, K be a compact subset of M, and $\mathcal{O}$ be an open neighbourhood of K in M. There exists $\hat{M} = \hat{M}(L, t_0, K, \mathcal{O}) > 0$ such that, for every critical viscosity subsolution of class $C^{1,1}_{loc}$ such that the mapping $x \mapsto d_x u$ is L-Lipschitz on $\mathcal{O}$, we have*

$$c(H) - H(x, d_x u) \leq \hat{M} \left\{ h_t(x,x) + c(H)t \right\}^{\frac{1}{2}}, \quad \forall t \geq t_0, \quad \forall x \in K.$$

Proof. Let $x \in K$ be such that $H(x, d_x u) < c(H)$. For every $y \in M$, we set $\epsilon(y) := c(H) - H(y, d_y u)$. Since the mapping $y \mapsto \epsilon(y)$ is continuous, there exists a constant C_L such that

$$\epsilon(y) \leq C_L, \quad \forall y \in K.$$

Moreover, since $y \mapsto d_y u$ is L-Lipschitz on $\mathcal{O}$, there exists $K_L > 0$ such that the mapping $y \mapsto H(y, d_y u)$ is K_L-Lipschitz on $\mathcal{O}$ and $\frac{C_L}{2K_L} \leq \operatorname{dist}(K, \mathcal{O}^c)$. Hence $B\left(x, \frac{\epsilon(x)}{2K_L}\right) \subset \mathcal{O}$ and so we have

$$H(y, d_y u) \leq c(H) - \frac{\epsilon(x)}{2}, \quad \forall y \in B\left(x, \frac{\epsilon(x)}{2K_L}\right)$$

Since K is compact and L is uniformly superlinear in the fibers, there exists an upper bound A for $\|w\|_y$ over the set of (y,w) with $y \in K$ such that the corresponding periodic orbit in TM minimizes $h_t(y,y)$ for some $t \geq t_0$ (this follows directly from the proof of Proposition 6.2.17 in the Appendix). Let $\gamma : [0,t] \to M$ be such that $\gamma(0) = \gamma(t) = x$ and satisfying

$$h_t(x,x) = \int_0^t L(\gamma(s), \dot{\gamma}(s)) ds.$$

Since $\|\dot{\gamma}(s)\|_{\gamma(s)}$ is always bounded by A, we have

$$\gamma(s) \in B\left(x, \frac{\epsilon(x)}{2K_L}\right), \quad \forall |s| \leq \frac{\epsilon(x)}{2AK_L} =: s_0(x).$$

Thus we have

$$u(\gamma(s_0(x))) - u(x) \leq \int_0^{s_0(x)} L(\gamma(s), \dot\gamma(s)ds + \left(c(H) - \frac{\epsilon(x)}{2}\right) s_0(x).$$

Moreover since u is a critical viscosity subsolution, if $t \geq s_0(x)$ we have

$$u(x) - u(\gamma(s_0(x))) \leq \int_{s_0(x)}^{t} L(\gamma(s), \dot\gamma(s)ds + c(H)(t - s_0(x)).$$

In consequence, we obtain

$$\begin{aligned} 0 &\leq \int_0^t L(\gamma(s), \dot\gamma(s)ds + c(H)t - \frac{\epsilon(x)}{2} s_0(x) \\ &= \{h_t(x,x) + c(H)t\} - \frac{\epsilon(x)}{2} s_0(x). \end{aligned}$$

Therefore we have

$$\{h_t(x,x) + c(H)t\} \geq \frac{\epsilon(x)}{2} s_0(x) = \frac{\epsilon^2(x)}{4AK_L},$$

which means that, as long as $s_0(x) \leq t_0$,

$$c(H) - H(x, d_x u) \leq 2\sqrt{AK_L}\, \{h_t(x,x) + c(H)t\}^{\frac{1}{2}}, \quad \forall t \geq t_0.$$

Thus, in order to conclude, we need to have $s_0(x) = \frac{\epsilon(x)}{2AK_L} \leq t_0$ for all $x \in K$, which is the case if

$$\frac{C_L}{2AK_L} \leq t_0 \;\Leftrightarrow\; AK_L \geq \frac{C_L}{2t_0}.$$

Then we conclude

$$c(H) - H(x, d_x u) \leq \max\left\{ 2\sqrt{AK_L}, \sqrt{2\frac{C_L}{t_0}} \right\} \{h_t(x,x) + c(H)t\}^{\frac{1}{2}}, \quad \forall t \geq t_0, \quad \forall x \in K.$$

□

Returning to the proof of Theorem 4.1.4, we notice that, without loss of generality, we can assume that we work in a compact set K which is included in a relatively compact open subset $\mathcal{O}$ of $\mathbb{R}^n$ with $n = \dim M$. So, let us denote by $L = L(\mathcal{O})$ the uniform Lipschitz constant given by Theorem 4.3.1. Fix from now a pair of $C^{1,1}_{loc}$ critical subsolutions $v_1, v_2 : M \to \mathbb{R}$ such that $d_x v_1, d_x v_2$ are L-Lipschitz on $\mathcal{O}$. We notice that, by strict convexity of the Hamiltonian, there exists a constant $\beta > 0$ such that for every $x \in \mathcal{O}$,

$$H\left(x, \frac{p_1+p_2}{2}\right) \leq \frac{H(x,p_1)+H(x,p_2)}{2} - \beta|p_1-p_2|^2, \quad \forall p_1, p_2 \in T_x^*M.$$

Hence, we have for every $x \in \mathcal{O}$,

$$H\left(x, \frac{d_xv_1 + d_xv_2}{2}\right) \leq c(H) - \beta|d_xv_1 - d_xv_2|^2.$$

By Lemma 4.4.1, we deduce that

$$|d_xv_1 - d_xv_2|^2 \leq \frac{\hat{M}}{\beta}\left\{h_t(x,x) + c(H)t\right\}^{\frac{1}{2}}, \quad \forall t \geq \bar{T} - \delta(\bar{x}), \quad \forall x \in K. \tag{4.4.4}$$

Let $(x,v),(y,w) \in \tilde{A}_i$. From Lemma 4.2.3, there is a C^1 path $(x(\cdot),v(\cdot)) : [0,1] \to N$ in $\tilde{B}_i$ from (x,v) to (y,w) with length less than $C_i|(x,v)-(y,w)|$. Since $d_xv_1 = \mathcal{L}(x,v)$ and $d_yv_1 = \mathcal{L}(y,w)$, there is a constant $D > 0$ such that for every $s \in [0,1]$,

$$\begin{aligned} |(x(s),v(s)) - (x,v)| &\leq C_i|(x,v)-(y,w)| \\ &\leq C_i\sqrt{1+D^2}|x-y|. \end{aligned}$$

Hence, by (4.4.3), we have for every $s \in [0,1]$,

$$\Psi(x(s),v(s)) \leq M_iC_i^{k-1}\left(1+D^2\right)^{\frac{k-1}{2}}|x-y|^{k-1}.$$

By (4.4.2), this means that for every $s \in [0,1]$,

$$h_{\rho(x(s),v(s))}(x(s),x(s)) + c(H)\rho(x(s),v(s)) \leq (1 + c(H) + J)D_i|x-y|^{\frac{k-1}{2}},$$

for a certain constant D_i. By (4.4.4) we finally deduce that there exists some constant $E_i > 0$ such that, for every $s \in [0,1]$,

$$\left|d_{x(s)}v_1 - d_{x(s)}v_2\right| \leq E_i|x-y|^{\frac{k-1}{8}}.$$

Integrating along the path, we obtain

$$|(v_1-v_2)(x) - (v_1-v_2)(y)| \leq CE_i|x-y|^{\frac{k-1}{8}+1}.$$

Finally, from Lemma 4.2.1, we deduce that

$$d_M(x,y) \leq CE_i|x-y|^{\frac{k-1}{8}+1},$$

for every $x,y \in \pi(\tilde{A}_i) \cap \mathcal{O}$. As a consequence, we deduce easily that, if $\frac{k-1}{8}+1 \leq n$ (that is $k \geq 8n-7$), the semi-metric space $A_{\bar{x}}$ has vanishing one-dimensional Hausdorff measure. Using a countable family of sets $A_{\bar{x}}$ to cover $\mathcal{A}_M^p$, we conclude easily that the one-dimensional Hausdorff dimension of $(\mathcal{A}_M^p, d_M)$ vanishes as soon as $k \geq 8n-7$. Finally, the fact that the α-dimensional Hausdorff measure of $(\mathcal{A}_M^p, d_M)$ vanishes whenever $\alpha \in (0,1]$ is such that $\alpha(\frac{k-1}{8}+1) \geq \dim M$ follows from the same arguments as the proof above.

4.4.4 A general result

We observe that most of the results proved above can be seen as particular cases of the following general principle.

Theorem 4.4.2. *Assume that* $\dim M \geq 3$*,* H *is of class* C^2*, and that there are* $r > 0$*,* $k', l \in \mathbb{N}$ *and a function* $G : TM \to \mathbb{R}$ *of class* $C^{k',1}$ *which satisfies the following properties:*

1. $G(x, v) \equiv 0$ *on* $\tilde{\mathcal{A}}$*;*
2. $\{m_r(x)\}^l \leq G(x, v)$ *for all* $(x, v) \in TM$*, where* $m_r(x) := \inf_{t \geq r} \{h_t(x, x) + c(H)t\}$*;*
3. $k' \geq 4l(\dim M - 1) - 1$*.*

Then $(\mathcal{A}_M, d_M)$ *has vanishing one-dimensional Hausdorff measure. Moreover, if* $\alpha \in (0, 1]$ *is such that* $\alpha(\frac{k'+1}{4l} + 1) \geq \dim M$ *then* $(\mathcal{A}_M, d_M)$ *has vanishing* α*-dimensional Hausdorff measure. In particular, if* G *is* C^∞*, then* $(\mathcal{A}_M, d_M)$ *has zero Hausdorff dimension.*

Proof. As in the proof of Proposition 4.1.4 we can assume that we work in a compact set K which is included in a relatively compact open subset $\mathcal{O}$ of $\mathbb{R}^n$ with $n = \dim M$. Let $L = L(\mathcal{O})$ be the uniform Lipschitz constant given by Theorem 4.3.1, and $v_1, v_2 : M \to \mathbb{R}$ be two $C^{1,1}_{loc}$ critical subsolutions such that $d_x v_1, d_x v_2$ are L-Lipschitz on $\mathcal{O}$. By lemma 4.4.1, there exists $\hat{M} > 0$ such that

$$c(H) - H(x, d_x u_1) \leq \hat{M} \{m_r(x)\}^{\frac{1}{2}}, \quad c(H) - H(x, d_x u_2) \leq \hat{M} \{m_r(x)\}^{\frac{1}{2}}, \quad \forall x \in K.$$

Arguing now as in the proof of Theorem 4.1.4, we get

$$|d_x v_1 - d_x v_2|^2 \leq \frac{\hat{M}}{\beta} \{m_r(x)\}^{\frac{1}{2}} \leq \frac{\hat{M}}{\beta} G(x, v)^{\frac{1}{2l}}, \quad \forall x \in K.$$

From Lemma 4.2.2, we can decompose the set $\tilde{A}$ as

$$\tilde{A} = \cup_{i \in \mathbb{N}} \tilde{A}_i.$$

Let $i \geq 0$ be fixed. Since G is a nonnegative $C^{k',1}$ function vanishing on $\tilde{A}$, by (4.2.1), we know that

$$0 \leq G(x, v) \leq M_i \left(|x - y|^2 + |v, w|^2\right)^{\frac{k'+1}{2}}, \quad \forall (y, w) \in \tilde{A}_i, \quad \forall (x, v) \in \tilde{B}_i.$$

As in the proof of Theorem 4.1.4, we deduce easily that there is a constant $N_i > 0$ such that

$$d_M(x, y) \leq N_i |x - y|^{\frac{k'+1}{4l} + 1},$$

for every $x, y \in \pi(\tilde{A}_i) \cap \mathcal{O}$. We now conclude as in the proofs of Theorems 4.1.2 and 4.1.4. ☐

Remark 4.4.3. It can be shown that for every compact subset $K \subset M$, there is a constant $C_K > 0$ such that

$$h(x,x) \leq C_K d(x,\mathcal{A})^2, \quad \forall x \in K,$$

where $d(x,\mathcal{A})$ denotes the Riemannian distance from x to the set $\mathcal{A}$ (which is assumed to be nonempty). Therefore, from Theorem 4.4.2, we deduce that if there are $l \in \mathbb{N}$ and a function $G : M \to \mathbb{R}$ of class $C^{k',1}$ with $k' \geq 2l(\dim M - 1) - 1$ such that

$$d(x,\mathcal{A})^l \leq G(x), \quad \forall x \in M,$$

then $(\mathcal{A}_M, d_M)$ has vanishing one-dimensional Hausdorff measure.

4.5 Proof of Theorem 4.1.5

By Theorems 4.1.2 and 4.1.4 we know that $(\mathcal{A}^0_M \cup \mathcal{A}^p_M, d_M)$ has zero Hausdorff dimension. Theorem 4.1.5 will follow from the fact that $\mathcal{A}_M \setminus (\mathcal{A}^0_M \cup \mathcal{A}^p_M)$ is a finite set.

We recall that the Aubry set $\tilde{\mathcal{A}} \subset TM$ is defined as the set of $(x,v) \in TM$ such that $x \in \mathcal{A}$ and v is the unique $v \in T_xM$ such that $d_x u = \frac{\partial L}{\partial v}(x,v)$ for any critical viscosity subsolution. This set is invariant under the Euler-Lagrange flow ϕ^L_t. For every $x \in \mathcal{A}$, we denote by $\mathcal{O}(x)$ the projection on $\mathcal{A}$ of the orbit of ϕ^L_t which passes through x. We observe that the following simple fact holds:

Lemma 4.5.1. *If $x, y \in \mathcal{A}$ and $\overline{\mathcal{O}(x)} \cap \overline{\mathcal{O}(y)} \neq \emptyset$, then $d_M(x,y) = 0$.*

Let us define

$$\mathcal{C}_0 := \{x \in \mathcal{A} \mid \overline{\mathcal{O}(x)} \cap \mathcal{A}_0\}, \quad \mathcal{C}_p := \{x \in \mathcal{A} \mid \overline{\mathcal{O}(x)} \cap \mathcal{A}_p\}.$$

Thus, if $x \in \mathcal{C}_0 \cup \mathcal{C}_p$, by Lemma 4.5.1 the Mather distance between x and $\mathcal{A}^0 \cup \mathcal{A}^p$ is zero, and we have done.

Let us now define $\mathcal{C} := \mathcal{A} \setminus (\mathcal{C}_0 \cup \mathcal{C}_p)$, and let $(\mathcal{C}_M, d_M)$ be the quotiented metric space. To conclude the proof, we show that this set consists of a finite number of points.

Let u be a $C^{1,1}$ critical subsolution (which existence is provided by [20]), and let X be the Lipschitz vector field uniquely defined by the relation

$$\mathcal{L}(x, X(x)) := (x, d_x u),$$

where $\mathcal{L}$ denotes the Legendre transform. Its flow extends on the whole manifold the flow considered above on $\mathcal{A}$. We fix $x \in \mathcal{C}$. Then $\overline{\mathcal{O}(x)}$ is a non-empty, compact, invariant set which contains a non-trivial minimal set for the flow of X (see [115, Chapter 1]). By [101], we know that there exists at most a finite number of such non-trivial minimal sets. Therefore, again by Lemma 4.5.1, $(\mathcal{C}_M, d_M)$ consists only in a finite number of points.

4.6 Applications in Dynamics

4.6.1 Preliminary results

Throughout this section, M is assumed to be compact. As before, $H : T^*M \to \mathbb{R}$ is a Hamiltonian of class at least C^2 satisfying the three usual conditions (H1)-(H3), and L is the Tonelli Lagrangian which is associated to it by Fenchel's duality. We denote by $\mathcal{SS}$ the set of critical viscosity subsolutions and by $\mathcal{S}_-$ the set of weak KAM solutions. Hence $\mathcal{S}_- \subset \mathcal{SS}$. If $u : M \to \mathbb{R}$ is a critical viscosity subsolution, we recall, see [65], that the set $\tilde{\mathcal{I}}(u)$ is the subset of TM defined by

$$\tilde{\mathcal{I}}(u) = \{(x,v) \in TM \mid t \mapsto \gamma(t) := \pi\phi_t(x,v) \text{ is } (u,L,c(H))\text{-calibrated on } (-\infty,+\infty)\},$$

that is for all $t_1 < t_2 \in \mathbb{R}$,

$$u(\gamma(t_2)) - u(\gamma(t_1)) = \int_{t_1}^{t_2} L(\gamma(t),\dot\gamma(t))\,dt + c(H)(t_2 - t_1).$$

The following properties of $\tilde{\mathcal{I}}(u)$ are shown in [65].

Theorem 4.6.1. *The set $\tilde{\mathcal{I}}(u)$ is invariant under the Euler-Lagrange flow ϕ_t^L. Moreover, if $(x,v) \in \tilde{\mathcal{I}}(u)$, then d_xu exists, and we have*

$$d_xu = \frac{\partial L}{\partial v}(x,v) \quad \text{and} \quad H(x,d_xu) = c(H).$$

It follows that:

1) *The restriction $\pi|_{\tilde{\mathcal{I}}(u)}$ of the projection is injective; therefore, if we set $\mathcal{I}(u) = \pi(\tilde{\mathcal{I}}(u))$, then $\tilde{\mathcal{I}}(u)$ is a continuous graph over $\mathcal{I}(u)$.*

2) *The map $x \mapsto d_xu$ is continuous on $\mathcal{I}(u)$.*

Moreover the following results holds (see [65] or [63, Théorème 1]).

Theorem 4.6.2. *1) Two weak KAM solutions that coincide on $\mathcal{A}$ are equal everywhere. 2) For every $u \in \mathcal{SS}$, there is a unique weak KAM solution $u_- : M \to \mathbb{R}$ such that $u_- = u$ on $\mathcal{A}$; moreover, the two functions u and u_- are also equal on $\mathcal{I}(u)$.*

The Aubry set $\tilde{\mathcal{A}}$ is given by

$$\tilde{\mathcal{A}} := \bigcap_{u \in \mathcal{SS}} \tilde{\mathcal{I}}(u) = \bigcap_{u \in \mathcal{S}_-} \tilde{\mathcal{I}}(u)$$

(the equivalence between the definition with $\mathcal{SS}$ and the one with $\mathcal{S}_-$ can be easily shown from the results of [65]). The projected Aubry set $\mathcal{A}$ is simply the image $\pi(\tilde{\mathcal{A}})$. We also have

$$\mathcal{A} := \bigcap_{u\in\mathcal{SS}} \mathcal{I}(u) = \bigcap_{u\in\mathcal{S}_-} \mathcal{I}(u).$$

The Mañé set $\tilde{\mathcal{N}}$ is given by

$$\tilde{\mathcal{N}} := \bigcup_{u\in\mathcal{SS}} \tilde{\mathcal{I}}(u) = \bigcup_{u\in\mathcal{S}_-} \tilde{\mathcal{I}}(u).$$

Both $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{N}}$ are compact subsets of TM invariant under the Euler-Lagrange flow ϕ_t^L of L.

Theorem 4.6.3 (Mañé). *Each point of the invariant set $\tilde{\mathcal{A}}$ is chain-recurrent for the restriction $\phi_t^L|_{\tilde{\mathcal{A}}}$. Moreover, the invariant set $\tilde{\mathcal{N}}$ is chain-transitive for the restriction $\phi_t^L|_{\tilde{\mathcal{N}}}$.*

Corollary 4.6.4. *The restriction $\phi_t^L|_{\tilde{\mathcal{A}}}$ to the invariant subset $\tilde{\mathcal{A}}$ is chain-transitive if and only if $\tilde{\mathcal{A}}$ is connected.*

Proof. This is an easy well-known result in the theory of Dynamical Systems: Suppose θ_t, $t \in \mathbb{R}$, is a flow on the compact metric space X. If every point of X is chain-recurrent for θ_t, then θ_t is chain-transitive if and only if X is connected. □

We give now the general relationship between uniqueness of weak KAM solutions and the quotient Mather set.

Proposition 4.6.5. *The two following statements are equivalent:*

1) *Any two weak KAM solutions differ by a constant.*

2) *The Mather quotient (A_M, d_M) is trivial, i.e. is reduced to one point.*

Moreover, if anyone of these conditions is true, then $\tilde{\mathcal{A}} = \tilde{\mathcal{N}}$, and therefore $\tilde{\mathcal{A}}$ is connected and the restriction of the Euler-Lagrange flow ϕ_t^L to $\tilde{\mathcal{A}}$ is chain-transitive.

Proof. For every fixed $x \in M$, the function $y \mapsto h(x, y)$ is a weak KAM solution. Therefore if we assume that any two weak KAM solutions differ by a constant, then for $x_1, x_2 \in M$ we can find a constant C_{x_1,x_2} such that

$$\forall y \in M, \quad h(x_1, y) = C_{x_1,x_2} + h(x_2, y).$$

If $x_2 \in \mathcal{A}$, then $h(x_2, x_2) = 0$, therefore evaluating the equality above for $y = x_2$, we obtain $C_{x_1,x_2} = h(x_1, x_2)$. Substituting in the equality and evaluating we conclude

$$\forall x_1 \in M, \quad \forall x_2 \in \mathcal{A}, \quad h(x_1, x_1) = h(x_1, x_2) + h(x_2, x_1).$$

This implies

$$\forall x_1, x_2 \in \mathcal{A}, \quad h(x_1, x_2) + h(x_2, x_1) = 0.$$

Which means that $d_M(x_1, x_2) = 0$, for every $x_1, x_2 \in \mathcal{A}$.

To prove the converse, let us recall that for every critical subsolution u, we have

$$\forall x, y \in M, \quad u(y) - u(x) \leq h(x, y).$$

Therefore applying this for a pair $u_1, u_2 \in \mathcal{SS}$, we obtain

$$\begin{aligned} \forall x, y \in M, \quad & u_1(y) - u_1(x) \leq h(x, y), \\ & u_2(x) - u_2(y) \leq h(y, x). \end{aligned}$$

Adding and rearranging, we obtain

$$\forall x, y \in M, \quad (u_1 - u_2)(y) - (u_1 - u_2)(x) \leq h(x, y) + h(y, x).$$

Since the right hand side is symmetric in x, y, we obtain

$$\forall x, y \in M, \quad |(u_1 - u_2)(y) - (u_1 - u_2)(x)| \leq h(x, y) + h(y, x).$$

If we assume that 2) is true, this implies that $u_1 - u_2$ is constant c on the projected Aubry set $\mathcal{A}$, or $u_1 = u_2 + c$ on $\mathcal{A}$. If u_1, u_2 are in fact weak KAM solutions then we must have $u_1 = u_2 + c$ on M, because any two solutions equal on the Aubry set are equal everywhere, see 2) of Theorem 4.6.2.

It remains to show the last statement. Notice that if $u_1, u_2 \in \mathcal{SS}$ differ by a constant then $\tilde{\mathcal{I}}(u_1) = \tilde{\mathcal{I}}(u_2)$. Therefore if any two elements in $\mathcal{S}_-$ differ by a constant, then

$$\tilde{\mathcal{A}} = \tilde{\mathcal{I}}(u) = \tilde{\mathcal{N}},$$

where u is any element in $\mathcal{S}_-$. But, by Mañé's Theorem 4.6.3, the invariant set $\tilde{\mathcal{N}}$ is chain-transitive for the flow ϕ_t, hence it is connected by Corollary 4.6.4. □

We now denote by X_L the Euler-Lagrange vector field of L, that is the vector field on TM that generates ϕ_t^L. We recall that an important property of X_L is that

$$\forall (x, v) \in TM, \quad T\pi(X_L(x, v)) = v,$$

where $T\pi : T(TM) \to TM$ denotes the canonical projection.
Here is a last ingredient that we will have to use.

Proposition 4.6.6 (Lyapunov Property). *Suppose $u_1, u_2 \in \mathcal{SS}$. The function $(u_1 - u_2) \circ \pi$ is non-decreasing along any orbit of the Euler Lagrange flow ϕ_t^L contained in $\tilde{\mathcal{I}}(u_2)$. If we assume u_1 is differentiable at $x \in \mathcal{I}(u_2)$, and $(x, v) \in \tilde{\mathcal{I}}(u_2)$, then, using that u_2 is differentiable on $\mathcal{I}(u_2)$, we obtain*

$$X_L \cdot [(u_1 - u_2) \circ \pi](x, v) = d_x u_1(v) - d_x u_2(v) \leq 0.$$

Moreover, the inequality above is an equality, if and only if $d_x u_1 = d_x u_2$. In that case $H(x, d_x u_1) = H(x, d_x u_2) = c(H)$.

Proof. If $(x, v) \in \tilde{\mathcal{I}}(u_2)$ then $t \mapsto \pi\phi_t(x, v)$ is $(u_2, L, c(H))$-calibrated, hence

$$\forall t_1 \leq t_2, \quad u_2 \circ \pi(\phi_{t_2}(x, v)) - u_2 \circ \pi(\phi_{t_1}(x, v)) = \int_{t_1}^{t_2} L(\phi_s(x, v))\, ds + c(H)(t_2 - t_1).$$

Since $u_1 \in \mathcal{SS}$, we get

$$\forall t_1 \leq t_2, \quad u_1 \circ \pi(\phi_{t_2}(x, v)) - u_1 \circ \pi(\phi_{t_1}(x, v)) \leq \int_{t_1}^{t_2} L(\phi_s(x, v))\, ds + c(H)(t_2 - t_1).$$

Combining these two facts, we conclude

$$\forall t_1 \leq t_2, \quad u_1 \circ \pi(\phi_{t_2}(x, v)) - u_1 \circ \pi(\phi_{t_1}(x, v)) \leq u_2 \circ \pi(\phi_{t_2}(x, v)) - u_2 \circ \pi(\phi_{t_1}(x, v)).$$

This implies

$$\forall t_1 \leq t_2, \quad (u_1 - u_2) \circ \pi(\phi_{t_2}(x, v)) \leq (u_1 - u_2) \circ \pi(\phi_{t_1}(x, v)).$$

Recall that u_2 is differentiable at every $x \in \mathcal{I}(u_2)$. Thus, if also $d_x u_1$ exists, if $(x, v) \in \tilde{\mathcal{I}}(u_2)$ we obtain

$$X_L \cdot [(u_1 - u_2) \circ \pi](x, v) \leq 0.$$

We remark that $X_L \cdot [(u_1 - u_2) \circ \pi](x, v) = d_x(u_1 - u_2)(T\pi \circ X_L(x, v))$. Since $T\pi \circ X_L(x, v) = v$, we obtain

$$X_L \cdot [(u_1 - u_2) \circ \pi](x, v) = d_x u_1(v) - d_x u_2(v) \leq 0.$$

If the last inequality is an equality, we have $d_x u_1(v) = d_x u_2(v))$. Since $(x, v) \in \tilde{\mathcal{I}}(u_2)$, we have $d_x u_2 = \frac{\partial L}{\partial v}(x, v)$ and $H(x, d_x u_2) = c(H)$, therefore the Fenchel inequality yields the equality

$$d_x u_2(v) = L(x, v) + H(x, d_x u_2) = L(x, v) + c(H).$$

Since $u_1 \in \mathcal{SS}$, we know that $H(x, d_x u_1) \leq c(H)$. The previous equality, using the Fenchel inequality $d_x u_1(v) \leq L(x, v) + H(x, d_x u_1)$, and the fact that $d_x u_1(v) = d_x u_2(v)$, implies

$$H(x, d_x u_1) = c(H) \quad \text{and} \quad d_x u_1(v) = L(x, v) + H(x, d_x u_1).$$

This means that we have equality in the Fenchel inequality $d_x u_1(v) \leq L(x, v) + H(x, d_x u_1)$, we therefore conclude that $d_x u_1 = \frac{\partial L}{\partial v}(x, v)$, but the right hand side of this last equality is $d_x u_2$. □

4.6.2 Strong Mather condition

Definition 4.6.7. We will say that the the Tonelli Lagrangian L on M satisfies the *strong Mather condition* if for every pair $u_1, u_2 \in \mathcal{S}_-$, the image $(u_1 - u_2)(\mathcal{A}) \subset \mathbb{R}$ is of Lebesgue measure 0.

Notice that by part 2) of Theorem 4.6.2, if L satisfies the strong Mather condition, then for every pair of critical sub-solutions u_1, u_2, the image $(u_1 - u_2)(\mathcal{A}) \subset \mathbb{R}$ is also of Lebesgue measure 0. By the results proved in this chapter, we get:

Theorem 4.6.8. *Let L be a Tonelli Lagrangian on the compact manifold M. It satisfies the strong Mather condition in any one of the following cases:*

(1) The dimension of M is 1 or 2.

(2) The dimension of M is 3, and $\tilde{\mathcal{A}}$ contains no fixed point of the Euler-Lagrange flow.

(3) The dimension of M is 3, and L is of class $C^{3,1}$.

(4) The Lagrangian is of class $C^{k,1}$, with $k \geq 2 \dim M - 3$, and every point of $\tilde{\mathcal{A}}$ is fixed under the Euler-Lagrange flow ϕ_t^L.

(5) The Lagrangian is of class C^k, with $k \geq 8 \dim M - 7$, and each point of $\tilde{\mathcal{A}}$ either is fixed under the Euler-Lagrange flow ϕ_t^L or its orbit in the Aubry set is periodic with (strictly) positive period.

Lemma 4.6.9. *Suppose that L is a Tonelli Lagrangian L on the compact manifold M that satisfies the strong Mather condition. For every $u \in \mathcal{SS}$, the set of points in $\tilde{\mathcal{I}}(u)$ which are chain-recurrent for the restriction $\phi_t^L|_{\tilde{\mathcal{I}}(u)}$ of the Euler-Lagrange flow is precisely the Aubry set $\tilde{\mathcal{A}}$.*

Proof. First of all, we recall that, from Theorem 4.6.3, each point of $\mathcal{A}$ is chain-recurrent for the restriction $\phi_t^L|_{\tilde{\mathcal{A}}}$. By [69, Theorem 1.5], we can find a C^1 critical viscosity subsolution $u_1 : M \to \mathbb{R}$ which is strict outside $\mathcal{A}$, *i.e.* for every $x \notin \mathcal{A}$ we have $H(x, d_x u_1) < c(H)$. We define θ on TM by $\theta = (u_1 - u) \circ \pi$. By Proposition 4.6.1, we know that at each point (x, v) of $\tilde{\mathcal{I}}(u)$ the derivative of θ exists and depends continuously on $(x, v) \in \tilde{\mathcal{I}}(u)$. By Proposition 4.6.6, at each point of (x, v) of $\tilde{\mathcal{I}}(u)$, we have

$$X_L \cdot \theta(x, v) = d_x u_1(v) - d_x u(v)) \leq 0,$$

with the last inequality an equality if and only if $d_x u_1 = d_x u$, and this implies $H(x, d_x u_1) = c(H)$. Since u_1 is strict outside $\mathcal{A}$, we conclude that $X_L \cdot \theta < 0$ on $\tilde{\mathcal{I}}(u) \setminus \tilde{\mathcal{A}}$. Suppose that $(x_0, v_0) \in \tilde{\mathcal{I}}(u) \setminus \tilde{\mathcal{A}}$. By invariance of both $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{I}}(u)$, every point on the orbit $\phi_t^L(x_0, v_0), t \in \mathbb{R}$ is also contained in $\tilde{\mathcal{I}}(u) \setminus \tilde{\mathcal{A}}$, therefore $t \mapsto c(t) := \theta(\phi_t(x_0, v_0))$ is (strictly) decreasing , and so we have $c(1) < c(0)$. Observe now that $\theta(\tilde{\mathcal{A}}) = (u_1 - u)(\mathcal{A})$ has measure 0 by the strong Mather condition, therefore we can find $c \in]c(1), c(0)[\setminus \theta(\tilde{\mathcal{A}})$. By what we have seen, the directional derivative $X_L \cdot \theta$ is < 0 at every point of the level set $L_c = \{(x, v) \in \tilde{\mathcal{I}}(u) \mid \theta(x, v) = c\}$. Since θ is everywhere non-increasing on the orbits of ϕ_t^L and $X_L \cdot \theta < 0$ on L_c, we get

$$\forall t > 0, \quad \forall (x, v) \in L_c, \quad \theta(\phi_t(x, v)) < c.$$

Consider the compact set $K_c = \{(x, v) \in \tilde{\mathcal{I}}(u) \mid \theta(x, v) \le c\}$. Using again that θ is non-increasing on the orbits of $\phi_t^L|_{\tilde{\mathcal{I}}(u)}$, we have

$$\forall t \ge 0, \quad \phi_t^L(K_c) \subset K_c \quad \text{and} \quad \phi_t^L(K_c \setminus L_c) \subset K_c \setminus L_c.$$

Using what we obtained above on L_c, we conclude that

$$\forall t > 0, \quad \phi_t^L(K_c) \subset K_c \setminus L_c.$$

We fix now some metric on $\tilde{\mathcal{I}}(u)$ defining its topology. We consider then the compact set $\phi_1^L(K_c)$. It is contained in the open set $K_c \setminus L_c = \{(x, v) \in \tilde{\mathcal{I}}(u) \mid \theta(x, v) < c\}$. We can therefore find $\epsilon > 0$ such that the ϵ neighborhood $V_\epsilon(\phi_1(K_c))$ of $\phi_1^L(K_c)$ in $\tilde{\mathcal{I}}(u)$ is also contained in K_c. Since for $t \ge 1$, we have $\phi_{t-1}^L(K_c) \subset K_c$, and therefore $\phi_t^L(K_c) \subset \phi_1(K_c)$, it follows that

$$V_\epsilon\left(\bigcup_{t \ge 1} \phi_t^L(K_c)\right) \subset K_c.$$

It is know easy to conclude that every ϵ-pseudo orbit for $\phi_t^L|_{\tilde{\mathcal{I}}(u)}$ that starts in K_c remains in K_c. Since $\theta(\phi_1^L(x_0, v_0)) = c(1) < c < c(0) = \theta(x_0, v_0)$, no α-pseudo orbit starting at (x_0, v_0) can return to (x_0, v_0), for $\alpha \le \epsilon$ such that the ball of center $\phi_1^L(x_0, v_0)$ and radius α, in $\tilde{\mathcal{I}}(u)$, is contained in K_c. Therefore (x_0, v_0) cannot be chain recurrent. □

Theorem 4.6.10. *Let L be a Tonelli Lagrangian on the compact manifold M. If L satisfies the strong Mather condition, then the following statements are equivalent:*

(1) The Aubry set $\tilde{\mathcal{A}}$, or its projection $\mathcal{A}$, is connected.

(2) The Aubry set $\tilde{\mathcal{A}}$ is chain-transitive for the restriction of the Euler-Lagrange flow $\phi_t^L|_{\tilde{\mathcal{A}}}$.

(3) Any two weak KAM solutions differ by a constant.

(4) The Aubry set $\tilde{\mathcal{A}}$ is equal to the Mañé set $\tilde{\mathcal{N}}$.

(5) There exists $u \in \mathcal{SS}$ such that $\tilde{\mathcal{I}}(u)$ is chain-recurrent for the restriction $\phi_t|_{\tilde{\mathcal{I}}(u)}$ of the Euler-Lagrange flow.

Proof. From Corollary 4.6.4, we know that (1) and (2) are equivalent.

If (1) is true then for $u_1, u_2 \in \mathcal{S}_-$, the image $u_1 - u_2(\mathcal{A})$ is a sub-interval of $\mathbb{R}$, but by the strong Mather condition, it is also of Lebesgue measure 0, therefore $u_1 - u_2$ is constant. Hence (1) implies (3).

If (3) is true then (4) follows from Proposition 4.6.5.

Suppose now that (4) is true. Since for every $u \in \mathcal{SS}$, we have $\tilde{\mathcal{A}} \subset \tilde{\mathcal{I}}(u) \subset \tilde{\mathcal{N}}$, we obtain $\tilde{\mathcal{I}}(u) = \tilde{\mathcal{N}}$. But $\tilde{\mathcal{N}}$ is chain-transitive for the restriction $\phi_t^L|_{\tilde{\mathcal{N}}}$. Hence (4) implies (5).

If (5) is true for some $u \in \mathcal{SS}$, then every point of $\tilde{\mathcal{I}}(u)$ is chain-recurrent for the restriction $\phi_t^L|_{\tilde{\mathcal{I}}(u)}$. Lemma 4.6.9 then implies that $\tilde{\mathcal{A}} = \tilde{\mathcal{I}}(u)$, and we therefore satisfy (2). □

4.6.3 Mañé Lagrangians

We give know an application to the Mañé example associated to a vector field. Suppose M is a compact Riemannian manifold, where the metric g is of class C^∞. If X is a C^k vector field on M, with $k \geq 2$, we define the Lagrangian $L_X : TM \to \mathbb{R}$ by

$$L_X(x, v) = \frac{1}{2}\|v - X(x)\|_x^2,$$

where as usual $\|v - X(x)\|_x^2 = g_x(v, v)$. We will call L_X the Mañé Lagrangian of X, see the Appendix in [99]. The following proposition gives the obvious properties of L_X.

Proposition 4.6.11. *Let L_X the Mañé Lagrangian of the $C^k, k \geq 2$, vector field X on the compact Riemannian manifold M. We have*

$$\frac{\partial L_X}{\partial v}(x, v) = g_x(v - X(x), \cdot).$$

Its associated Hamitonian $H_X : T^*M \to \mathbb{R}$ is given by

$$H_X(x, p) = \frac{1}{2}\|p\|_x^2 + p(X(x)).$$

The constant functions are solutions of the Hamilton-Jacobi equation

$$H_X(x, d_x u) = 0.$$

Therefore, we obtain $c(H) = 0$. Moreover, we have

$$\tilde{\mathcal{I}}(0) = \mathrm{Graph}(X) = \{(x, X(x)) \mid x \in M\}.$$

If we call ϕ_t the Euler-Lagrange flow of L_X on TM, then for every $x \in M$, and every $t \in \mathbb{R}$, we have $\phi_t(x, X(x)) = (\gamma_x^X(t), \dot\gamma_x^X(t))$, where γ_x^X is the solution of the vector field X which is equal to x for $t = 0$. In particular, the restriction $\phi_t|_{\tilde{\mathcal{I}}(0)}$ of the Euler-Lagrange flow to $\tilde{\mathcal{I}}(0) = \mathrm{Graph}(X)$ is conjugated (by $\pi|_{\tilde{\mathcal{I}}(0)}$) to the flow of X on M.

Proof. The computation of $\frac{\partial L_X}{\partial v}$ is easy. For H_X, we recall that $H_X(x, p) = p(v_p) - L(x, v_p)$, where $v_p \in T_xM$ is defined by $p = \frac{\partial L_X}{\partial v}(x, v_p)$. Solving for v_p, and substituting yields the result.

If u is a constant function then $d_xu = 0$ everywhere, and obviously $H_X(x, d_xu) = 0$. The fact that $c(H) = 0$ follows, since $c(H)$ is the only value c for which there exists a viscosity solution of the Hamilton-Jacobi equation $H(x, d_xu) = c$.

Let us define u_0 as the null function on M. Suppose now that $\gamma : (-\infty, +\infty) \to M$ is a solution of X (by compactness of M solutions of X are defined for all time). We have $d_{\gamma(t)}u_0(\dot\gamma(t)) = 0$, and $H_X(\gamma(t), d_{\gamma(t)}u_0) = 0$; moreover, since $\dot\gamma(t) = X(\gamma(t))$, we also get $L_X(\gamma(t), \dot\gamma(t)) = 0$. It follows that

$$d_{\gamma(t)}u_0(\dot\gamma(t)) = L_X(\gamma(t), \dot\gamma(t)) + H_X(\gamma(t), d_{\gamma(t)}u_0) = L_X(\gamma(t), \dot\gamma(t)).$$

By integration, we see that γ is $(u_0, L_X, 0)$-calibrated, therefore it is an extremal. Hence we get $\phi_t(\gamma(0), \dot\gamma(0)) = (\gamma(t), \dot\gamma(t))$, and $(\gamma(0), \dot\gamma(0)) \in \tilde{\mathcal{I}}(u_0)$. But $\dot\gamma(0) = X(\gamma(0))$, and $\gamma(0)$ can be an arbitrary point of M. This implies $\mathrm{Graph}(X) \subset \tilde{\mathcal{I}}(u_0)$. This finishes the proof because we know that $\tilde{\mathcal{I}}(u_0)$ is a graph on a part of the base M. □

Lemma 4.6.12. *Let $L_X : TM \to \mathbb{R}$ be the Mañé Lagrangian associated to the $C^k, k \geq 2$, vector field X on the compact connected manifold M. Assume that L_X satisfies the strong Mather condition, then we have:*

(1) The projected Aubry set $\mathcal{A}$ is the set of chain-recurrent points of the flow of X on M.

(2) The constants are the only weak KAM solutions if and only every point of M is chain-recurrent under the flow of X.

Proof. To prove (1), we apply Lemma 4.6.9 to obtain that the Aubry set $\tilde{\mathcal{A}}$ is equal to set of points in $\tilde{\mathcal{I}}(0)) = \mathrm{Graph}(X)$ which are chain-recurrent for the restriction $\phi_t|_{\mathrm{Graph}(X)}$. But from Proposition 4.6.11, then projection $\pi|_{\mathrm{Graph}(X)}$ conjugates $\phi_t|_{\mathrm{Graph}(X)}$ to the flow of X on M. It now suffices to observe that $\mathcal{A} = \pi(\tilde{\mathcal{A}})$.

We now prove (2). Suppose that every point of M is chain-recurrent for the flow of X. From what we have just seen $\mathcal{A} = M$. Therefore property (1) of Theorem 4.6.10 holds. Therefore by property (3) of that same theorem, we have uniqueness up to constants of weak KAM solutions, but the constants are weak KAM solutions. To prove the converse, assume that the constants are the only weak KAM solutions. This implies that property (3) of Theorem 4.6.10 holds. Therefore by property (4) of that same theorem $\tilde{\mathcal{A}} = \tilde{\mathcal{N}}$.

But $\tilde{\mathcal{I}}(0)) = \text{Graph}(X)$ is squeezed between $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{N}}$. Therefore $\tilde{\mathcal{A}} = \text{Graph}(X)$. Taking images by the projection π we conclude that $\mathcal{A} = M$. By part (1) of the present lemma, every point of M is chain-recurrent for the flow of X on M. □

Combining this last lemma and Theorem 4.6.8 completes the proof of Theorem 4.1.7.

Chapter 5

DiPerna-Lions theory for SDE

5.1 Introduction and preliminary results

[1] Recent research activity has been devoted to study transport equations with rough coefficients, showing that a well-posedness result for the transport equation in a certain subclass of functions allows to prove existence and uniqueness of a flow for the associated ODE. The first result in this direction is due to DiPerna and P.-L.Lions [56], where the authors study the connection between the transport equation and the associated ODE $\dot\gamma = b(t,\gamma)$, showing that existence and uniqueness for the transport equation is equivalent to a sort of well-posedness of the ODE which says, roughly speaking, that the ODE has a unique solution for $\mathscr{L}^d$-almost every initial condition (here and in the sequel, $\mathscr{L}^d$ denotes the Lebesgue measure in $\mathbb{R}^d$). In that paper they also show that the transport equation $\partial_t u + \sum_i b_i \partial_i u = c$ is well-posed in L^∞ if $b = (b_1, \ldots, b_n)$ is Sobolev and satisfies suitable global conditions (including L^∞-bounds on the spatial divergence), which yields the well-posedness of the ODE.

In [4] (see also [5]), using a slightly different philosophy, Ambrosio studied the connection between the continuity equations $\partial_t u + \sum_i \partial_i(b_i u) = c$ and the ODE $\dot\gamma = b(t,\gamma)$. This different approach allows him to develop the general theory of the so-called Regular Lagrangian Flows (see [5, Remark 31] for a detailed comparison with the DiPerna-Lions axiomatization), which relates existence and uniqueness for the continuity equation with well-posedness of the ODE, without assuming any regularity on the vector field b. Indeed, since the transport equation is in a conservative form, it has a meaning in the sense of distributions even when b is only L^∞_{loc} and u is L^1_{loc}. Thus, a general theory is developed in [4] under very general hypotheses, showing as in [56] that existence and uniqueness

[1]This chapter is based on the work in [77].

for the continuity equation is equivalent to a sort of well-posedness of the ODE. After having proved this, in [4] the well-posedness of the continuity equations in L^∞ is proved in the case of vector fields with BV regularity whose distributional divergence belongs to L^∞ (for other similar results on the well-posedness of the transport/continuity equation, see also [47, 48, 90, 83]).
Our aim is to develop a stochastic counterpart of this theory: in our setting the continuity equation becomes the Fokker-Planck equation, while the ODE becomes an SDE.
Let us consider the following SDE

$$\begin{cases} dX = b(t,X)\,dt + \sigma(t,X)\,dB(t) \\ X(0) = x, \end{cases} \tag{5.1.1}$$

where $b : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T]\times\mathbb{R}^d \to \mathscr{L}(\mathbb{R}^r,\mathbb{R}^d)$ are bounded (here $\mathscr{L}(\mathbb{R}^r,\mathbb{R}^d)$ denotes the vector space of linear maps from $\mathbb{R}^r$ to $\mathbb{R}^d$) and B is an r-dimensional Brownian motion on a probability space $(\Omega,\mathcal{A},\mathbb{P})$. We want to study the existence and uniqueness of martingale solutions for this equation. Let us define $a(t,x) := \sigma(t,x)\sigma^*(t,x)$ (that is $a_{ij} := \sum_k \sigma_{ik}\sigma_{jk}$). We consider the so called Fokker-Planck equation

$$\begin{cases} \partial_t\mu_t + \sum_i \partial_i(b_i\mu_t) - \frac{1}{2}\sum_{ij}\partial_{ij}(a_{ij}\mu_t) = 0 & \text{in } [0,T]\times\mathbb{R}^d, \\ \mu_0 = \bar{\mu} & \text{in } \mathbb{R}^d. \end{cases} \tag{5.1.2}$$

We recall that, for a (possibly signed) measure $\mu = \mu(t,x) = \mu_t(x)$, being a solution of (5.1.2) simply means that

$$\frac{d}{dt}\int_{\mathbb{R}^d}\varphi(x)\,d\mu_t(x) = \int_{\mathbb{R}^d}\Big[\sum_i b_i(t,x)\partial_i\varphi(x) + \frac{1}{2}\sum_{ij} a_{ij}(t,x)\partial_{ij}\varphi(x)\Big]\,d\mu_t(x) \quad \forall\varphi\in C_c^\infty(\mathbb{R}^d). \tag{5.1.3}$$

in the distributional sense on $[0,T]$, and the initial condition means that μ_t w^*-converges to $\bar{\mu}$ (i.e. converges in the duality with $C_c(\mathbb{R}^d)$) as $t \to 0$. We observe that, since the equation (5.1.2) is in divergence form, it makes sense without any regularity assumption on a and b, provided that

$$\int_0^T \int_A \big(|b(t,x)| + |a(t,x)|\big)\,d|\mu_t|(x)\,dt < +\infty \quad \forall A \subset\subset \mathbb{R}^d$$

(here and in the sequel, $|\mu_t|$ denotes the total variation of μ_t). Since b and a will always be assumed to be bounded, in the definition of measure-valued solution of the PDE we assume that

$$\int_0^T |\mu_t|(A)\,dt < +\infty \quad \forall A \subset\subset \mathbb{R}^d, \tag{5.1.4}$$

so that (5.1.2) surely makes sense. However, if μ_t is singular with respect to the Lebesgue measure $\mathscr{L}^d$, then the products $b(t,\cdot)\mu_t$ and $a(t,\cdot)\mu_t$ are sensitive to modification of $b(t,\cdot)$ and $a(t,\cdot)$ in $\mathscr{L}^d$-negligible sets. Since in the case of singular measures the coefficients a and b will be assumed to be continuous, while in the case of coefficients in L^∞ the measures will be assumed to be absolutely continuous, (5.1.2) will always make sense.
Recall also that it is not restrictive to consider only solutions $t \mapsto \mu_t$ of the Fokker-Planck equation that are w^*-continuous on $[0,T]$, i.e. continuous in the duality with $C_c(\mathbb{R}^d)$ (see Lemma 5.2.1). Thus, we can assume that μ_t is defined for all t and even at the endpoints of $[0,T]$.

For simplicity of notation, we define

$$L_t := \sum_i b_i(t,\cdot)\partial_i + \frac{1}{2}\sum_{ij} a_{ij}(t,\cdot)\partial_{ij}.$$

In this way the PDE can be written as

$$\partial_t \mu_t = L_t^* \mu_t \quad \text{in } [0,T]\times\mathbb{R}^d,$$

where L_t^* denotes the (formal) adjoint of L_t in $L^2(\mathbb{R}^d)$. Using Itô's formula it is simple to check that, if $X(t,x,\omega) \in L^2(\Omega, C([0,T],\mathbb{R}^d))$ is a family of solutions of (5.1.1), measurable in (t,x,ω), then the measure μ_t defined by

$$\int f(x)\,d\mu_t(x) := \int \mathbb{E}[f(X(t,x,\omega))]\,d\overline{\mu}(x) \quad \forall f \in C_c(\mathbb{R}^d)$$

is a solution of (5.1.2) with $\mu_0 = \overline{\mu}$ (see also Lemma 5.2.4).
We define $\Gamma_T := C([0,T],\mathbb{R}^d)$, and $e_t : \Gamma_T \to \mathbb{R}^d$, $e_t(\gamma) := \gamma(t)$. Let us recall the Stroock-Varadhan definition of martingale solutions:

Definition 5.1.1. A measure $\nu_{x,s}$ on Γ_T is a martingale solution of (5.1.1) starting from x at time s if:

1. $\nu_{x,s}(\{\gamma \mid \gamma(s) = x\}) = 1$;
2. for any $\varphi \in C_c^\infty(\mathbb{R}^d)$, the stochastic process on Γ_T

$$\varphi(\gamma(t)) - \int_s^t L_u\varphi(\gamma(u))\,du$$

   is a $\nu_{x,s}$-martingale after time s with respect to the canonical filtration.

We will say that the martingale problem is well-posed if, for any $(s,x) \in \mathbb{R}^d$, we have existence and uniqueness of martingale solutions.

In the sequel, we will deal with families $\{\nu_x\}_{x\in\mathbb{R}^d}$ of probability measures that are measurable with respect to x according to the following standard definition.

Definition 5.1.2. We say that a family of probability measures on a probability space $(\Omega, \mathcal{A})$ $\{\nu_x\}_{x\in\mathbb{R}^d}$ is measurable if, for any $A \in \mathcal{A}$, the real valued map $x \mapsto \nu_x(A)$ is measurable.

5.1.1 Plan of the chapter

• **The theory of Stochastic Lagrangian Flows**
In the first part, we develop a general theory (independent of specific regularity or ellipticity assumptions), which roughly speaking allows to deduce existence, uniqueness and stability of martingale solutions for $\mathscr{L}^d$-almost every initial condition x whenever existence and uniqueness is known at the PDE level in the L^∞-setting (and, conversely, if existence and uniqueness of martingale solutions is known for $\mathscr{L}^d$-a.e. initial condition, then existence and uniqueness for the PDE in the L^∞-setting holds).

More precisely, in Section 5.2 we study how uniqueness of the SDE is related to that of the PDE. In Paragraph 5.2.1 we prove a representation formula for solutions of the PDE, which shows that they can always be seen as a superposition of solutions of the SDE also when standard existence results for martingale solutions of SDE do not apply. In particular, assuming only the boundedness of the coefficients, we will show that, whenever we have existence of a solution of the PDE starting from μ_0, there exists at least one martingale solution of the SDE for μ_0-a.e. initial condition x.

In Section 5.3 we introduce the main object of our study, what we call Stochastic Lagrangian Flow. In Paragraph 5.3.1 we state and prove our main result regarding the existence and uniqueness of Stochastic Lagrangian Flows, showing that these flows exist and are unique whenever the PDE is well-posed in the L^∞-setting. We also prove a stability result, and we show that Stochastic Lagrangian Flows satisfy the Chapman-Kolmogorov equation. Moreover, in Paragraph 5.3.2 we investigate the relation between our result and its deterministic counterpart and, applying our stability result, we deduce a vanishing viscosity theorem for Ambrosio's Regular Lagrangian Flows.

• **The Fokker-Planck equation**
In the second part we study by purely PDE methods the well-posedness of the Fokker-Planck equation in two extreme (with respect to the regularity imposed in time, or in space) situations: in the first one, assuming uniform ellipticity of the coefficients and Lipschitz regularity in time, we are able to prove existence and uniqueness in the L^2-settings assuming no regularity in space, but only suitable divergence bounds (see Theorem 5.4.3). This result, together with Proposition 5.4.4, directly implies the following theorem (here

and in the sequel, $\mathcal{S}_+(\mathbb{R}^d)$ denotes the set of symmetric and non-negative definite $d \times d$ matrices).

Theorem 5.1.3. *Let us assume that* $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ *and* $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ *are bounded functions such that:*

1. $\sum_j \partial_j a_{ij} \in L^\infty([0,T] \times \mathbb{R}^d)$ *for* $i = 1, \dots, d$,
2. $\partial_t a_{ij} \in L^\infty([0,T] \times \mathbb{R}^d)$ *for* $i, j = 1, \dots, d$;
3. $(\sum_i \partial_i b_i - \frac{1}{2} \sum_{ij} \partial_{ij} a_{ij})^- \in L^\infty([0,T] \times \mathbb{R}^d)$;
4. $\langle \xi, a(t,x)\xi \rangle \geq \alpha |\xi|^2 \ \forall (t,x) \in [0,T] \times \mathbb{R}^d$, *for a certain* $\alpha > 0$;
5. $\frac{a}{1+|x|^2} \in L^2([0,T] \times \mathbb{R}^d)$, $\frac{b}{1+|x|} \in L^2([0,T] \times \mathbb{R}^d)$.

Then there exist a unique solution of (5.1.2) in $\mathscr{L}_+$, *where*

$$\mathscr{L}_+ := \left\{ u \in L^\infty([0,T], L^1_+(\mathbb{R}^d)) \cap L^\infty([0,T], L^\infty_+(\mathbb{R}^d)) \mid u \in C([0,T], w^* - L^\infty(\mathbb{R}^d)) \right\},$$

and L^1_+ *and* L^∞_+ *denote the convex subsets of* L^1 *and* L^∞ *consisting of non-negative functions.*

In the second case, a does not depend on the space variables, but it can be degenerate and it is allowed to depend on t even in a measurable way. Since a can also be identically 0, we need to assume BV regularity on the vector field b, and so we can prove:

Theorem 5.1.4. *Let us assume that* $a : [0,T] \to \mathcal{S}_+(\mathbb{R}^d)$ *and* $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ *are bounded functions such that:*

1. $b \in L^1([0,T], BV_{loc}(\mathbb{R}^d, \mathbb{R}^d))$, $\sum_i \partial_i b_i \in L^1_{loc}([0,T] \times \mathbb{R}^d)$;
2. $(\sum_i \partial_i b_i)^- \in L^1([0,T], L^\infty(\mathbb{R}^d))$.

Then there exist a unique solution of (5.1.2) in $\mathscr{L}_+$.

This theorem is a direct consequence of Theorem 5.4.12. Other existence and uniqueness results for the Fokker-Planck equation, which are in some sense intermediate with respect the two extreme ones stated above, have been proved in a recent paper of LeBris and P.-L.Lions [91]. As in our case, in that paper the authors are interested in the well-posedness of the Fokker-Planck equation as a tool to deduce existence and uniqueness results at the SDE level (see also [92]). In particular, in [91, Section 4] the authors give a list of interesting situations in the modelization of polymeric fluids when SDEs with irregular drift b and dispersion matrix σ arise (see also [88] and the references therein for other existence and uniqueness results for non-smooth SDEs).

• **Conclusions**
In Section 5.5 we apply the theory developed in Paragraph 5.3.1 to obtain, in the cases considered above, the generic well-posedness of the associated SDE.

Finally, in the last section we generalize an important uniqueness result of Stroock and Varadhan (see Theorem 5.2.2 and the remarks at the end of Theorem 5.5.4).

5.2 SDE-PDE uniqueness

In this section we study the main relations between the SDE and the PDE. The main result is a general representation formula for solutions of the PDE (Theorem 5.2.7) which allows to relate uniqueness of the SDE to that of the PDE (Lemma 5.2.3).

As we already said in the introduction, here and in the sequel b and a are always assumed to be bounded. Let us recall the following result on the time regularity of $t \mapsto \mu_t$ (see also [5, Remark 3] or [11, Lemma 8.1.2]):

Lemma 5.2.1. *Up to modification of μ_t in a negligible set of times, $t \mapsto \mu_t$ is w^*-continuous on $[0,T]$. Moreover, if $|\mu_t|(\mathbb{R}^d) \leq C$ for any $t \in [0,T]$, then $t \mapsto \mu_t$ is narrowly continuous.*

Proof. By (5.1.3), for any $\varphi \in C_c^\infty(\mathbb{R}^d)$

$$\frac{d}{dt}\int_{\mathbb{R}^d} \varphi(x)\, d\mu_t(x) \in L^1([0,T]),$$

and therefore, for a given φ, the map $t \mapsto \langle \mu_t, \varphi\rangle$ has a unique uniformly continuous representative in $[0,T]$. By a simple density argument, we can find a representative $\tilde\mu_t$ independent of $\varphi \in C_c^\infty(\mathbb{R}^d)$ such that $t \mapsto \langle \tilde\mu_t, \varphi\rangle$ is uniformly continuous on $[0,T]$. Together with (5.1.4), this implies that $t \mapsto \langle \tilde\mu_t, \varphi\rangle$ is uniformly continuous for any $\varphi \in C_c(\mathbb{R}^d)$. If moreover $|\mu_t|(\mathbb{R}^d) \leq C$ for any $t \in [0,T]$, then $t \mapsto \langle \tilde\mu_t, \varphi\rangle$ is uniformly continuous for any $\varphi \in C_b(\mathbb{R}^d)$. □

We also recall the following important theorem of Stroock and Varadhan (for a proof, see [125, Theorem 6.2.3]):

Theorem 5.2.2. *Assume that for any $(s,x) \in [0,T]\times\mathbb{R}^d$, for any $\nu_{x,s}$ and $\tilde\nu_{x,s}$ martingale solutions of (5.1.1) starting from x at time s, one has*

$$(e_t)_\#\nu_{x,s} = (e_t)_\#\tilde\nu_{x,s} \quad \forall t \in [s,T].$$

Then the martingale solution of (5.1.1) starting from any $(s,x) \in [0,T]\times\mathbb{R}^d$ is unique.

We start studying how the uniqueness of (5.1.1) is related to that of (5.1.2).

Lemma 5.2.3. *Let $A \subset \mathbb{R}^d$ be a Borel set. The following two properties are equivalent:*

(a) Time-marginals of martingale solutions of the SDE are unique for any $x \in A$.

(b) Finite non-negative measure-valued solutions of the PDE are unique for any non-negative Radon measure μ_0 concentrated in A.

Proof. $(b) \Rightarrow (a)$: let us choose $\mu_0 = \delta_x$, with $x \in A$. Then, if ν_x and $\tilde{\nu}_x$ are two martingale solutions of the SDE, we get that $\mu_t := (e_t)_\# \nu_x$ and $\tilde{\mu}_t := (e_t)_\# \tilde{\nu}_x$ are two solutions of the PDE with $\mu_0 = \delta_x$ (see Lemma 5.2.4). This implies that $\mu_t = \tilde{\mu}_t$, that is

$$\langle \mu_t, \varphi \rangle = \int_{\Gamma_T} \varphi(\gamma(t))\, d\nu_x(\gamma) = \int_{\Gamma_T} \varphi(\gamma(t))\, d\tilde{\nu}_x(\gamma) = \langle \tilde{\mu}_t, \varphi \rangle \quad \forall \varphi \in C_c^\infty(\mathbb{R}^d),$$

that is $(e_t)_\# \nu_x = (e_t)_\# \tilde{\nu}_x$ (observe in particular that, if $A = \mathbb{R}^d$ and we have uniqueness for the PDE for any initial time $s \geq 0$, by Theorem 5.2.2 we get that $\nu_x = \tilde{\nu}_x$ for any $x \in \mathbb{R}^d$).
$(a) \Rightarrow (b)$: this implication follows by Theorem 5.2.7, which provides, for every finite non-negative measure-valued solutions of the PDE, the representation

$$\int_{\mathbb{R}^d} \varphi\, d\mu_t = \int_{\mathbb{R}^d \times \Gamma_T} \varphi(\gamma(t))\, d\nu_x(\gamma)\, d\mu_0(x), \tag{5.2.1}$$

where, for μ_0-a.e. x, ν_x is a martingale solution of SDE starting from x (at time 0). Therefore, by the uniqueness of $(e_t)_\# \nu_x$, we obtain that solutions of the PDE are unique. □

We now prove that, if ν_x is a martingale solution of the SDE starting from x (at time 0) for μ_0-a.e. x, the right hand side of (5.2.1) always defines a non-negative solution of the PDE. We recall that a locally finite measure in $\mathbb{R}^d$ is a possibly signed measure with locally finite total variation.

Lemma 5.2.4. *Let μ_0 be a locally finite measure on $\mathbb{R}^d$, and let $\{\nu_x\}_{x \in \mathbb{R}^d}$ be a measurable family of probability measures on Γ_T such that ν_x is a martingale solution of the SDE starting from x (at time 0) for $|\mu_0|$-a.e. x. Define on Γ_T the measure $\nu := \int_{\mathbb{R}^d} \nu_x\, d\mu_0(x)$, and assume that*

$$\int_0^T \int_{\mathbb{R}^d \times \Gamma_T} \chi_{B_R}(\gamma(t))\, d\nu_x(\gamma)\, d|\mu_0|(x)\, dt < +\infty \quad \forall R > 0. \tag{5.2.2}$$

Then the measure μ_t^ν on $\mathbb{R}^d$ defined by

$$\langle \mu_t^\nu, \varphi \rangle := \langle (e_t)_\# \nu, \varphi \rangle = \int_{\mathbb{R}^d \times \Gamma_T} \varphi(\gamma(t))\, d\nu_x(\gamma)\, d\mu_0(x) \quad \forall \varphi \in C_c^\infty(\mathbb{R}^d)$$

is a solution of the PDE.

Remark 5.2.5. Property 5.2.2 is trivially true if, for example, $|\mu_0|(\mathbb{R}^d) < +\infty$.

Proof. Let us first show that the map $t \mapsto \langle \mu_t^\nu, \varphi\rangle$ is absolutely continuous for any $\varphi \in C_c^\infty(\mathbb{R}^d)$. We recall that a real valued map $t \mapsto f(t)$ is said absolutely continuous if, for any $\varepsilon > 0$ there exists $\delta > 0$ such that, given any family of disjoint intervals $(s_k, t_k) \subset [0, T]$, the following implication holds:

$$\sum_k |t_k - s_k| \le \delta \quad \Rightarrow \quad \sum_k |f(t_k) - f(s_k)| \le \varepsilon.$$

Take $R > 0$ such that $\operatorname{supp}(\varphi) \subset B_R$, and let $I = \cup_{k=1}^n (s_k, t_k)$ be a subset of $[0, T]$ with (s_k, t_k) disjoint and such that $|t_k - s_k| \le 1$. For μ_0-a.e. x, by the definition of martingale solution we have

$$\int_{\Gamma_T} \varphi(\gamma(t_k))\, d\nu_x(\gamma) - \int_{\Gamma_T} \varphi(\gamma(s_k))\, d\nu_x(\gamma) = \int_{s_k}^{t_k} \int_{\Gamma_T} L_t\varphi(\gamma(t))\, d\nu_x(\gamma)\, dt$$
$$= \int_{s_k}^{t_k} \int_{\Gamma_T} \sum_i b_i(t, \gamma(t))\partial_i\varphi(\gamma(t))\, d\nu_x(\gamma)\, dt + \frac{1}{2}\int_{s_k}^{t_k} \int_{\Gamma_T} \sum_{ij} a_{ij}(t, \gamma(t))\partial_{ij}\varphi(\gamma(t))\, d\nu_x(\gamma)\, dt$$

and so, integrating with respect to μ_0, we obtain

$$|\langle \mu_{t_k}^\nu, \varphi\rangle - \langle \mu_{s_k}^\nu, \varphi\rangle| \le \|\varphi\|_{C^2}\Big[\|b\|_\infty + \frac{1}{2}\|a\|_\infty\Big] \int_{s_k}^{t_k} \int_{\mathbb{R}^d\times\Gamma_T} \chi_{B_R}(\gamma(t))\, d\nu_x(\gamma)\, d|\mu_0|(x)\, dt.$$

Thus

$$\sum_{k=1}^n |\langle \mu_{t_k}^\nu, \varphi\rangle - \langle \mu_{s_k}^\nu, \varphi\rangle| \le \|\varphi\|_{C^2}\Big[\|b\|_\infty + \frac{1}{2}\|a\|_\infty\Big] \sum_{k=1}^n \int_{s_k}^{t_k} \int_{\mathbb{R}^d\times\Gamma_T} \chi_{B_R}(\gamma(t))\, d\nu_x(\gamma)\, d|\mu_0|(x)\, dt,$$

which shows that the map $t \mapsto \langle \mu_t^\nu, \varphi\rangle$ is absolutely continuous thanks to (5.2.2) and the absolute continuity property of the integral. So, in order to conclude that μ_t^ν solves the PDE, it suffices to compute the time derivative of $t \mapsto \langle \mu_t^\nu, \varphi\rangle$, and, by the computation we made above, one simply gets

$$\frac{d}{dt}\langle \mu_t^\nu, \varphi\rangle = \int_{\mathbb{R}^d} \frac{d}{dt}\left(\int_{\Gamma_T} \varphi(\gamma(t))\, d\nu_x(\gamma)\right) d\mu_0(x)$$
$$= \int_{\mathbb{R}^d}\int_{\Gamma_T} L_t\varphi(\gamma(t))\, d\nu_x(\gamma)\, d\mu_0(x) = \langle \mu_t^\nu, L_t\varphi\rangle.$$

□

Remark 5.2.6. We observe that, by the definition of μ_t^ν, the following implications hold:

1. $\mu_0 \ge 0 \Rightarrow \forall t \ge 0$, $\mu_t^\nu \ge 0$ and $\mu_t^\nu(\mathbb{R}^d) = \mu_0(\mathbb{R}^d)$ (the total mass can also be infinite);
2. μ_0 signed $\Rightarrow \forall t \ge 0$, $|\mu_t^\nu|(\mathbb{R}^d) \le |\mu_0|(\mathbb{R}^d)$ (the total variation can also be infinite).

5.2.1 A representation formula for solutions of the PDE

We denote by $\mathcal{M}_+(\mathbb{R}^d)$ the set of non-negative finite measures on $\mathbb{R}^d$.

Theorem 5.2.7. *Let μ_t be a solution of the PDE such that $\mu_t \in \mathcal{M}_+(\mathbb{R}^d)$ for any $t \in [0,T]$, with $\mu_t(\mathbb{R}^d) \le C$ for any $t \in [0,T]$. Then there exists a measurable family of probability measures $\{\nu_x\}_{x\in\mathbb{R}^d}$ such that ν_x is a martingale solution of (5.1.1) starting from x (at time 0) for μ_0-a.e. x, and the following representation formula holds:*

$$\int_{\mathbb{R}^d} \varphi\, d\mu_t = \int_{\mathbb{R}^d\times\Gamma_T} \varphi(\gamma(t))\, d\nu_x(\gamma)\, d\mu_0(x). \tag{5.2.3}$$

By this theorem it follows that, whenever we have existence of a solution of the PDE starting from μ_0, there exists a martingale solution of the SDE for μ_0-a.e. initial condition x.

Proof. Up to a renormalization of μ_0, we can assume that $\mu_0(\mathbb{R}^d) = 1$.
Step 1: smoothing. Let $\rho : \mathbb{R}^d \to (0,+\infty)$ be a convolution kernel such that $|D^k\rho(x)| \le C_k|\rho(x)|$ for any $k \ge 1$ ($\rho(x) = Ce^{-\sqrt{1+|x|^2}}$, for instance). We consider the measures $\mu_t^\varepsilon := \mu_t * \rho_\varepsilon$. They are smooth solutions of the PDE

$$\partial_t\mu_t^\varepsilon + \sum_i \partial_i(b_i^\varepsilon\mu_t^\varepsilon) - \frac{1}{2}\sum_{ij}\partial_{ij}(a_{ij}^\varepsilon\mu_t^\varepsilon) = 0, \tag{5.2.4}$$

where $b_t^\varepsilon = b^\varepsilon(t,\cdot) := \frac{(b(t,\cdot)\mu_t)*\rho_\varepsilon}{\mu_t^\varepsilon}$, $a_t^\varepsilon = a^\varepsilon(t,\cdot) := \frac{(a(t,\cdot)\mu_t)*\rho_\varepsilon}{\mu_t^\varepsilon}$. Then it is immediate to see that

$$\|b_t^\varepsilon\|_\infty \le \|b_t\|_\infty, \quad \|a_t^\varepsilon\|_\infty \le \|a_t\|_\infty. \tag{5.2.5}$$

Since $|D^k\rho(x)| \le C_k|\rho(x)|$, it is simple to check that b^ε and a^ε are smooth and bounded together with all their spatial derivatives. By [125, Corollary 6.3.3], the martingale problem for a^ε and b^ε is well-posed (see Definition 5.1.1) and the family $\{\nu_x^\varepsilon\}_{x\in\mathbb{R}^d}$ of martingale solutions (starting at time 0) is measurable (see Definition 5.1.2). By (5.2.5) we can apply Lemma 5.2.4, which tells us that $\tilde\mu_t^\varepsilon := (e_t)_\# \int_{\mathbb{R}^d} \nu_x^\varepsilon\, d\mu_0^\varepsilon(x)$ is a finite measure which solves the smoothed PDE (5.2.4) with initial datum μ_0^ε. Then, since the solution of (5.2.4) is unique (Proposition 5.4.1), we obtain $\tilde\mu_t^\varepsilon = \mu_t^\varepsilon$, that is

$$\int_{\mathbb{R}^d} \varphi\, d\mu_t^\varepsilon = \int_{\mathbb{R}^d\times\Gamma_T} \varphi(\gamma(t))\, d\nu_x^\varepsilon(\gamma)\, d\mu_0^\varepsilon(x). \tag{5.2.6}$$

Step 2: tightness. It is clear that the measures $\mu_0^\varepsilon = \mu_0 * \rho_\varepsilon$ are tight. So, if we define $\nu^\varepsilon := \int_{\mathbb{R}^d} \nu_x^\varepsilon\, d\mu_0^\varepsilon$, we have

$$\lim_{R\to\infty} \sup_{0<\varepsilon<1} \nu^\varepsilon(\{|\gamma(0)| > R\}) = 0.$$

For any $\varphi \in C_c^\infty(\mathbb{R}^d)$, let us define $A_\varphi := \|\varphi\|_{C^2}\left[\|b\|_\infty + \frac{1}{2}\|a\|_\infty\right]$. Since for every $\varphi \in C_c^\infty(\mathbb{R}^d)$ and any $0 < \varepsilon < 1$

$$\varphi(\gamma(t)) - \int_0^t \Big(\sum_i b_i^\varepsilon(u,\gamma(u))\partial_i\varphi(\gamma(u)) + \frac{1}{2}\sum_{ij} a_{ij}^\varepsilon(u,\gamma(u))\partial_{ij}\varphi(\gamma(u))\Big)\,du$$

is a ν^ε-martingale with respect to the canonical filtration, by (5.2.5) we obtain that $\varphi(\gamma(t)) + A_\varphi t$ is a ν^ε-submartingale with respect to the canonical filtration. Thus [125, Theorem 1.4.6] can be applied, and the tightness of ν^ε follows.

Let ν be any limit point of ν^ε, and consider the disintegration of ν with respect to $\mu_0 = (e_0)_\#\nu$, i.e. $\nu = \int_{\mathbb{R}^d} \nu_x\, d\mu_0(x)$. Passing to the limit in (5.2.6), we get

$$\int_{\mathbb{R}^d} \varphi\, d\mu_t(x) = \int_{\mathbb{R}^d\times\Gamma_T} \varphi(\gamma(t))\, d\nu_x(\gamma)\, d\mu_0(x).$$

Step 3: ν_x is a martingale solution of the SDE for μ_0-a.e. x. Let $\varepsilon_n \to 0$ be a sequence such that ν is the weak limit of ν^{ε_n}. Let us fix a continuous function $f : \mathbb{R}^d \to \mathbb{R}$ with $0 \le f \le 1$, $s \in [0,T]$, and an $\mathcal{F}_s$-measurable continuous function $\Phi^s : \Gamma_T \to \mathbb{R}$ with $0 \le \Phi^s \le 1$, where $(\mathcal{F}_s)_{0\le s\le T}$ denotes the canonical filtration on Γ_T. We define

$$L_t^n := \sum_i b_i^{\varepsilon_n}(t,\cdot)\partial_i + \frac{1}{2}\sum_{ij} a_{ij}^{\varepsilon_n}(t,\cdot)\partial_{ij}.$$

Since each $\nu_x^{\varepsilon_n}$ is a martingale solution, we know that for any $t \in [s,T]$ and for any $\varphi \in C_c^\infty(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d\times\Gamma_T} \left[\varphi(\gamma(t)) - \int_0^t L_u^n\varphi(\gamma(u))\,du\right] \Phi^s(\gamma)\, d\nu_x^{\varepsilon_n}(\gamma) f(x)\, d\mu_0^{\varepsilon_n}(x)$$
$$= \int_{\mathbb{R}^d\times\Gamma_T} \left[\varphi(\gamma(s)) - \int_0^s L_u^n\varphi(\gamma(u))\,du\right] \Phi^s(\gamma)\, d\nu_x^{\varepsilon_n}(\gamma) f(x)\, d\mu_0^{\varepsilon_n}(x)$$

(see Definition 5.1.1), or equivalently

$$\int_{\mathbb{R}^d\times\Gamma_T} \left[\varphi(\gamma(t)) - \varphi(\gamma(s)) - \int_s^t L_u^n\varphi(\gamma(u))\,du\right] \Phi^s(\gamma)\, d\nu_x^{\varepsilon_n}(\gamma) f(x)\, d\mu_0^{\varepsilon_n}(x) = 0.$$

Let us take $\tilde b : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $\tilde a : [0,T]\times\mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ bounded and continuous, and define

$$\tilde L_t := \sum_i \tilde b_i(t,\cdot)\partial_i + \frac{1}{2}\sum_{ij} \tilde a_{ij}(t,\cdot)\partial_{ij},$$

$$\tilde{L}_t^n := \sum_i \tilde{b}_i^{\varepsilon_n}(t,\cdot)\partial_i + \frac{1}{2}\sum_{ij} \tilde{a}_{ij}^{\varepsilon_n}(t,\cdot)\partial_{ij},$$

where $\tilde{b}_i^{\varepsilon_n}$ and $\tilde{a}_{ij}^{\varepsilon_n}$ are defined analogously to $b_i^{\varepsilon_n}$ and $a_{ij}^{\varepsilon_n}$. Thus we can write

$$\begin{aligned}\int_{\mathbb{R}^d\times\Gamma_T} &\left[\varphi(\gamma(t)) - \varphi(\gamma(s)) - \int_s^t \tilde{L}_u^n \varphi(\gamma(u))\,du\right] \Phi^s(\gamma)\,d\nu_x^{\varepsilon_n}(\gamma) f(x)\,d\mu_0^{\varepsilon_n}(x) \\ &= \int_{\mathbb{R}^d\times\Gamma_T} \left[\int_s^t (L_u^n - \tilde{L}_u^n)\varphi(\gamma(u))\,du\right] \Phi^s(\gamma)\,d\nu_x^{\varepsilon_n}(\gamma) f(x)\,d\mu_0^{\varepsilon_n}(x).\end{aligned}$$

Then, recalling that $0 \le f \le 1$ and $0 \le \Phi^s \le 1$, we get

$$\begin{aligned}&\left|\int_{\mathbb{R}^d\times\Gamma_T} \left[\varphi(\gamma(t)) - \varphi(\gamma(s)) - \int_s^t \tilde{L}_u^n \varphi(\gamma(u))\,du\right] \Phi^s(\gamma)\,d\nu_x^{\varepsilon_n}(\gamma) f(x)\,d\mu_0^{\varepsilon_n}(x)\right| \\ &\quad\le \int_{\mathbb{R}^d\times\Gamma_T} \left[\int_s^t \left|(L_u^n - \tilde{L}_u^n)\varphi(\gamma(u))\right|\,du\right] \Phi^s(\gamma)\,d\nu_x^{\varepsilon_n}(\gamma) f(x)\,d\mu_0^{\varepsilon_n}(x) \\ &\quad\le \int_{\mathbb{R}^d\times\Gamma_T} \left[\int_s^t \left|(L_u^n - \tilde{L}_u^n)\varphi(\gamma(u))\right|\,du\right] d\nu_x^{\varepsilon_n}(\gamma)\,d\mu_0^{\varepsilon_n}(x) \\ &\quad= \int_s^t \int_{\mathbb{R}^d} \left|(L_u^n - \tilde{L}_u^n)\varphi(x)\right|\,d\mu_u^{\varepsilon_n}(x)\,du \\ &\quad\le \sum_i \int_s^t \int_{\mathbb{R}^d} \left|\left(\frac{(b_i(u,\cdot)\mu_u) * \rho_{\varepsilon_n}}{\mu_u^{\varepsilon_n}} - \frac{(\tilde{b}_i(u,\cdot)\mu_u) * \rho_{\varepsilon_n}}{\mu_u^{\varepsilon_n}}\right)\partial_i\varphi\right|(x)\,d\mu_u^{\varepsilon_n}(x)\,du \\ &\quad+ \frac{1}{2}\sum_{ij} \int_s^t \int_{\mathbb{R}^d} \left|\left(\frac{(a_{ij}(u,\cdot)\mu_u) * \rho_{\varepsilon_n}}{\mu_u^{\varepsilon_n}} - \frac{(\tilde{a}_{ij}(u,\cdot)\mu_u) * \rho_{\varepsilon_n}}{\mu_u^{\varepsilon_n}}\right)\partial_{ij}\varphi\right|(x)\,d\mu_u^{\varepsilon_n}(x)\,du \\ &\quad\le \sum_i \int_s^t \int_{\mathbb{R}^d} |b_i(u,\cdot) - \tilde{b}_i(u,\cdot)|(x)\partial_i\varphi * \rho_{\varepsilon_n}(x)\,d\mu_u(x)\,du \\ &\quad+ \frac{1}{2}\sum_{ij} \int_s^t \int_{\mathbb{R}^d} |a_{ij}(u,\cdot) - \tilde{a}_{ij}(u,\cdot)|(x)\partial_{ij}\varphi * \rho_{\varepsilon_n}(x)\,d\mu_u(x)\,du.\end{aligned}$$

Since $\tilde{a}$ and $\tilde{b}$ are continuous, $\tilde{a}^{\varepsilon_n}$ and $\tilde{b}^{\varepsilon_n}$ converge to $\tilde{a}$ and $\tilde{b}$ locally uniformly. So we can pass to the limit in the above equation as $n \to \infty$, obtaining

$$\begin{aligned}\left| \int_{\mathbb{R}^d \times \Gamma_T} \left[\varphi(\gamma(t)) - \varphi(\gamma(s)) - \int_s^t \tilde{L}_u \varphi(\gamma(u))\, du \right] \Phi^s(\gamma)\, d\nu_x(\gamma) f(x)\, d\mu_0(x) \right| \\ \leq \sum_i \int_s^t \int_{\mathbb{R}^d} |b_i(u,x) - \tilde{b}_i(u,x)| \partial_i \varphi(x)\, d\mu_u(x)\, du \\ + \frac{1}{2} \sum_{ij} \int_s^t \int_{\mathbb{R}^d} |a_{ij}(u,x) - \tilde{a}_{ij}(u,x)| \partial_{ij} \varphi(x)\, d\mu_u(x)\, du\end{aligned}$$

Choosing two sequences of continuous functions $(\tilde{b}^k)_{k\in\mathbb{N}}$ and $(\tilde{a}^k)_{k\in\mathbb{N}}$ converging respectively to b and a in $L^1([0,T] \times \mathbb{R}^d, \eta)$, with $\eta := \int_0^T \mu_t\, dt$, we finally obtain

$$\int_{\mathbb{R}^d \times \Gamma_T} \left[\varphi(\gamma(t)) - \varphi(\gamma(s)) - \int_s^t L_u \varphi(\gamma(u))\, du \right] \Phi^s(\gamma)\, d\nu_x(\gamma) f(x)\, d\mu_0(x) = 0,$$

that is

$$\begin{aligned}\int_{\mathbb{R}^d \times \Gamma_T} \left[\varphi(\gamma(t)) - \int_0^t L_u \varphi(\gamma(u))\, du \right] \Phi^s(\gamma)\, d\nu_x(\gamma) f(x)\, d\mu_0(x) \\ = \int_{\mathbb{R}^d \times \Gamma_T} \left[\varphi(\gamma(s)) - \int_0^s L_u \varphi(\gamma(u))\, du \right] \Phi^s(\gamma)\, d\nu_x(\gamma) f(x)\, d\mu_0(x).\end{aligned}$$

By the arbitrariness of f we get that, for any $0 \leq s \leq t \leq T$, and for any $\mathcal{F}_s$-measurable function Φ^s, we have

$$\begin{aligned}\int_{\Gamma_T} \left[\varphi(\gamma(t)) - \int_0^t L_u \varphi(\gamma(u))\, du \right] \Phi^s(\gamma)\, d\nu_x(\gamma) \\ = \int_{\Gamma_T} \left[\varphi(\gamma(s)) - \int_0^s L_u \varphi(\gamma(u))\, du \right] \Phi^s(\gamma)\, d\nu_x(\gamma) \quad \text{for } \mu_0\text{-a.e. } x.\end{aligned}$$

Letting Φ^s vary in a dense countable subset of $\mathcal{F}_s$-measurable functions, by approximations we deduce that, for any $0 \leq s \leq t \leq T$, for μ_0-a.e. x,

$$\begin{aligned}\int_{\Gamma_T} \left[\varphi(\gamma(t)) - \int_0^t L_u \varphi(\gamma(u))\, du \right] \Phi^s(\gamma)\, d\nu_x(\gamma) \\ = \int_{\Gamma_T} \left[\varphi(\gamma(s)) - \int_0^s L_u \varphi(\gamma(u))\, du \right] \Phi^s(\gamma)\, d\nu_x(\gamma)\end{aligned}$$

for any $\mathcal{F}_s$-measurable function Φ^s (here the μ_0-a.e. depends on s and t but not on Φ^s). Taking now $s, t \in [0,T] \cap \mathbb{Q}$, we deduce that, for μ_0-a.e. x,

$$\int_{\Gamma_T} \left[\varphi(\gamma(t)) - \int_0^t L_u\varphi(\gamma(u))\,du\right] \Phi^s(\gamma)\,d\nu_x(\gamma)$$
$$= \int_{\Gamma_T} \left[\varphi(\gamma(s)) - \int_0^s L_u\varphi(\gamma(u))\,du\right] \Phi^s(\gamma)\,d\nu_x(\gamma)$$

for any $s, t \in [0,T] \cap \mathbb{Q}$, for any $\mathcal{F}_s$-measurable function Φ^s. By the continuity of the above equality with respect to both s and t, and the continuity in time of the filtration $\mathcal{F}_s$, we conclude that ν_x is a martingale solution for μ_0-a.e. x. □

Remark 5.2.8. We observe that by (5.2.3) it follows that

$$\mu_t(\mathbb{R}^d) \le C \ \forall t \quad \Rightarrow \quad \mu_t(\mathbb{R}^d) = \mu_0(\mathbb{R}^d)$$

(this result can also be proved more directly using as test functions in (5.1.2) a suitable sequence $(\varphi_n)_{n\in\mathbb{N}} \subset C_c^\infty(\mathbb{R}^d)$, with $0 \le \varphi_n \le 1$ and $\varphi_n \nearrow 1$, and, even in the case when the measures μ_t are signed, under the assumption $|\mu_t|(\mathbb{R}^d) \le C$ one obtains the constancy of the map $t \mapsto \mu_t(\mathbb{R}^d)$).

5.3 Stochastic Lagrangian Flows

In this section we want to prove an existence and uniqueness result for martingale solutions which satisfy certain properties, in the spirit of the Regular Lagrangian Flows (RLF) introduced in [4].

Definition 5.3.1. Given a measure $\mu_0 = \rho_0\mathscr{L}^d \in \mathcal{M}_+(\mathbb{R}^d)$, with $\rho_0 \in L^\infty(\mathbb{R}^d)$, we say that a measurable family of probability measures $\{\nu_x\}_{x\in\mathbb{R}^d}$ on Γ_T is a μ_0-*Stochastic Lagrangian Flow* (μ_0-SLF) (starting at time 0), if:

1. for μ_0-a.e. x, ν_x is a martingale solution of the SDE starting from x (at time 0);
2. for any $t \in [0,T]$
$$\mu_t := (e_t)_\# \left(\int \nu_x\,d\mu_0(x)\right) \ll \mathscr{L}^d,$$
and, denoting $\mu_t = \rho_t\mathscr{L}^d$, we have $\rho_t \in L^\infty(\mathbb{R}^d)$ uniformly in t.

More in general, one can analogously define a μ_0-SLF starting at time s with $s \in (0,T)$ requiring that ν_x is a martingale solution of the SDE starting from x at time s.

Remark 5.3.2. If $\{\nu_x\}_{x\in\mathbb{R}^d}$ is a μ_0-SLF, then it is also a μ_0'-SLF for any $\mu_0' \in \mathcal{M}_+(\mathbb{R}^d)$ with $\mu_0' \le C\mu_0$. Indeed, this easily follows by the inequality

$$0 \le (e_t)_\# \int_{\mathbb{R}^d} \tilde{\nu}_x \, d\mu_0'(x) \le C (e_t)_\# \int_{\mathbb{R}^d} \tilde{\nu}_x \, d\mu_0(x).$$

5.3.1 Existence, uniqueness and stability of SLF

We denote by L^1_+ and L^∞_+ the convex subsets of L^1 and L^∞ consisting of non-negative functions, and, following [4], we define

$$\mathscr{L} := \left\{ u \in L^\infty([0,T], L^1(\mathbb{R}^d)) \cap L^\infty([0,T], L^\infty(\mathbb{R}^d)) \mid u \in C([0,T], w^* - L^\infty(\mathbb{R}^d)) \right\},$$

and

$$\mathscr{L}_+ := \left\{ u \in L^\infty([0,T], L^1_+(\mathbb{R}^d)) \cap L^\infty([0,T], L^\infty_+(\mathbb{R}^d)) \mid u \in C([0,T], w^* - L^\infty(\mathbb{R}^d)) \right\}.$$

Under an existence and uniqueness result for the PDE in the class $\mathscr{L}_+$, we prove existence and uniqueness of SLF.

Theorem 5.3.3 (Existence of SLF starting from a fixed measure). *Let us suppose that, for some initial datum $\mu_0 = \rho_0 \mathscr{L}^d \in \mathcal{M}_+(\mathbb{R}^d)$, with $\rho_0 \in L^\infty(\mathbb{R}^d)$, there exists a solution of the PDE in $\mathscr{L}_+$. Then there exists a μ_0-SLF.*

Proof. It suffices to apply Theorem 5.2.7 to the solution of the PDE in $\mathscr{L}_+$. □

Let us assume now that forward uniqueness for the PDE holds in the class $\mathscr{L}_+$ for any initial time, that is, for any $s \in [0,T]$, for any $\rho_s \in L^1_+(\mathbb{R}^d) \cap L^\infty_+(\mathbb{R}^d)$, if we denote by $\rho_t \mathscr{L}^d$ and $\tilde{\rho}_t \mathscr{L}^d$ two solutions of the PDE in the class $\mathscr{L}_+$ starting from $\rho_s \mathscr{L}^d$ at time s, then

$$\rho_t = \tilde{\rho}_t \quad \text{for any } t \in [s,T].$$

Before stating and proving our main theorem, we first introduce some notation that will be used also in the last section.

Let $\mathcal{B}$ be the Borel σ-algebra on $\Gamma_T = C([0,T], \mathbb{R}^d)$, and define the filtrations $\mathcal{F}_t := \sigma[e_s \mid 0 \le s \le t]$ and $\mathcal{F}^t := \sigma[e_s \mid t \le s \le T]$. Set $\mathcal{P}(\Gamma_T)$ the set of probability measures on Γ_T. Now, given $\nu \in \mathcal{P}(\Gamma_T)$, we denote by

$$\Gamma_T \ni \gamma \mapsto \nu^\gamma_{\mathcal{F}_t} \in \mathcal{P}(\Gamma_T)$$

a regular conditional probability distribution of ν given $\mathcal{F}_t$, that is a family of probability measures on $(\Gamma_T, \mathcal{B})$ indexed by γ such that:

- for each $B \in \mathcal{B}$, $\gamma \mapsto \nu^\gamma_{\mathcal{F}_t}(B)$ is $\mathcal{F}_t$-measurable;

-

$$\nu(A \cap B) = \int_A \nu^{\gamma}_{\mathcal{F}_t}(B)\, d\nu(\gamma) \quad \forall A \in \mathcal{F}_t,\ \forall B \in \mathcal{B}. \tag{5.3.1}$$

Since Γ_T is a Polish space and every σ-algebra $\mathcal{F}_t$ is finitely generated, such a function exists and is unique, up to ν-null sets. In particular, up to changing this function in a ν-null set, the following fact holds:

$$\nu^{\gamma}_{\mathcal{F}_t}(\{\tilde\gamma \mid \tilde\gamma(s) = \gamma(s)\ \forall s \in [0,t]\}) = 1 \quad \forall \gamma \in \Gamma_T. \tag{5.3.2}$$

Finally, given $0 \le t_1 \le \ldots \le t_n \le T$, we set $M^{t_1,\ldots,t_n} := \sigma[e_{t_1}, \ldots, e_{t_n}]$, and one can analogously define $\nu^{\gamma}_{M^{t_1,\ldots,t_n}}$. For $\nu^{\gamma}_{M^{t_1,\ldots,t_n}}$ an analogous of (5.3.2) holds:

$$\nu^{\gamma}_{M^{t_1,\ldots,t_n}}(\{\tilde\gamma \mid \tilde\gamma(t_i) = \gamma(t_i)\ \forall i = 1, \ldots, n\}) = 1 \quad \forall \gamma \in \Gamma_T. \tag{5.3.3}$$

If $\gamma(t_i) = x_i$ for $i = 1, \ldots, n$, then we will also use the notation $\nu^{\gamma}_{M^{t_1,\ldots,t_n}} = \nu^{x_1,\ldots,x_n}_{M^{t_1,\ldots,t_n}}$. By (5.3.1) one can check that $\int_{\Gamma_T} \nu^{\tilde\gamma}_{\mathcal{F}_{t_n}}\, d\nu^{\gamma}_{M^{t_1,\ldots,t_n}}(\tilde\gamma)$ is a regular conditional probability distribution of ν given $M^{t_1,\ldots,t_n}$, which implies by uniqueness that

$$\nu^{\gamma}_{M^{t_1,\ldots,t_n}} = \int_{\Gamma_T} \nu^{\tilde\gamma}_{\mathcal{F}_{t_n}}\, d\nu^{\gamma}_{M^{t_1,\ldots,t_n}}(\tilde\gamma) \quad \text{for } \nu\text{-a.e. } \gamma. \tag{5.3.4}$$

Theorem 5.3.4 (Uniqueness of SLF starting from a fixed measure). *Let us assume that forward uniqueness for the PDE holds in the class $\mathscr{L}_+$ for any initial time. Then, for any $\mu_0 = \rho_0 \mathscr{L}^d \in \mathcal{M}_+(\mathbb{R}^d)$, with $\rho_0 \in L^\infty(\mathbb{R}^d)$, the μ_0-SLF is uniquely determined μ_0-a.e. (in the sense that, if $\{\nu_x\}$ and $\{\tilde\nu_x\}$ are two μ_0-SLF, then $\nu_x = \tilde\nu_x$ for μ_0-a.e. x).*

Proof. Let $\{\nu_x\}$ and $\{\tilde\nu_x\}$ be two μ_0-SLF. Take now a function $\psi \in C_c(\mathbb{R}^d)$, with $\psi \ge 0$. By Remark 5.3.2, $\{\nu_x\}$ and $\{\tilde\nu_x\}$ are two $\psi\mu_0$-SLF. Thus, by Lemma 5.2.4 and the uniqueness of the PDE in $\mathscr{L}_+$, for any $\varphi \in C_c(\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d \times \Gamma_T} \varphi(e_t(\gamma))\, d\nu_x(\gamma)\psi(x)\, d\mu_0(x) = \int_{\mathbb{R}^d \times \Gamma_T} \varphi(e_t(\gamma))\, d\tilde\nu_x(\gamma)\psi(x)\, d\mu_0(x) \quad \forall t \in [0,T]. \tag{5.3.5}$$

This clearly implies that, for any $t \in [0,T]$,

$$(e_t)_\# \nu_x = (e_t)_\# \tilde\nu_x \quad \text{for } \mu_0\text{-a.e. } x.$$

We now want to use an analogous argument to deduce that, for any $0 < t_1 < t_2 < \ldots < t_n \le T$,

$$(e_{t_1}, \ldots, e_{t_n})_\# \nu_x = (e_{t_1}, \ldots, e_{t_n})_\# \tilde\nu_x \quad \text{for } \mu_0\text{-a.e. } x. \tag{5.3.6}$$

The idea is that, given a measure $\tilde{\mu}_s = \tilde{\rho}_s \mathscr{L}^d \in \mathcal{M}_+(\mathbb{R}^d)$, with $\tilde{\rho}_s \in L^\infty$, once we have a $\tilde{\mu}_s$-SLF starting at time s we can multiply $\tilde{\mu}_s$ by a function $\psi_s \in C_c(\mathbb{R}^d)$ with $\psi_s \geq 0$, and by Remark 5.3.2 our $\tilde{\mu}_s$-SLF is also a $\psi_s\tilde{\mu}_s$-SLF starting at time s. Using this argument n times at different times and the time marginals uniqueness, we will obtain (5.3.6).
Fix $0 < t_1 < \ldots < t_n \leq T$. Take $\psi_0 \geq 0$ with $\psi_0 \in C_c(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} \psi_0 \, d\mu_0 = 1$, and denote by $\mu_{t_1}^{\psi_0}$ the value at time t_1 of the (unique) solution in $\mathscr{L}_+$ of the PDE starting from $\psi_0\mu_0$ (which is induced both by $\{\nu_x\}$ and $\{\tilde{\nu}_x\}$ by uniqueness, see equation (5.3.5)). Let $\{\nu_{x,t_1}\}_{x\in\mathbb{R}^d}$ and $\{\tilde{\nu}_{x,t_1}\}_{x\in\mathbb{R}^d}$ be the families of probability measures on Γ_T given by the disintegration of

$$\nu^{\psi_0} := \int_{\mathbb{R}^d} \nu_x \psi_0(x) \, d\mu_0(x) \quad \text{and} \quad \tilde{\nu}^{\psi_0} := \int_{\mathbb{R}^d} \tilde{\nu}_x \psi_0(x) \, d\mu_0(x)$$

with respect to $\mu_{t_1}^{\psi_0} = (e_{t_1})_\# \nu^{\psi_0} = (e_{t_1})_\# \tilde{\nu}^{\psi_0}$, that is

$$\nu^{\psi_0} = \int_{\mathbb{R}^d} \nu_{x,t_1} \, d\mu_{t_1}^{\psi_0}(x), \quad \tilde{\nu}^{\psi_0} = \int_{\mathbb{R}^d} \tilde{\nu}_{x,t_1} \, d\mu_{t_1}^{\psi_0}(x). \tag{5.3.7}$$

It is easily seen that $\{\nu_{x,t_1}\}$ and $\{\tilde{\nu}_{x,t_1}\}$ are regular conditional probability distributions, given $M^{t_1} = \sigma[e_{t_1}]$, of ν^{ψ_0} and $\tilde{\nu}^{\psi_0}$ respectively (that is, with the notation introduced before, $\nu_{x,t_1} = (\nu^{\psi_0})^x_{M_{t_1}}$ and $\tilde{\nu}_{x,t_1} = (\tilde{\nu}^{\psi_0})^x_{M_{t_1}}$). Thus, looking at $\{\nu_{x,t_1}\}$ and $\{\tilde{\nu}_{x,t_1}\}$ as their restriction to $C([t_1,T],\mathbb{R}^d)$, $\{\nu_{x,t_1}\}$ and $\{\tilde{\nu}_{x,t_1}\}$ are $\mu_{t_1}^{\psi_0}$-SLF starting at time t_1. Indeed, by the stability of martingale solutions with respect to regular conditional probability (see [125, Chapter 6]), $\{\nu_{x,t_1}\}$ and $\{\tilde{\nu}_{x,t_1}\}$ are martingale solutions of the SDE starting from x at time t_1 for $\mu_{t_1}^{\psi_0}$-a.e. x (see also the remarks at the end of the proof of Proposition 5.6.1), while (ii) of Definition 5.3.1 is trivially true since $\{\nu_x\}$ and $\{\tilde{\nu}_x\}$ are $\psi_0\mu_0$-SLF. As before, since $\{\nu_{x,t_1}\}$ and $\{\tilde{\nu}_{x,t_1}\}$ are also $\psi_1\mu_{t_1}^{\psi_0}$-SLF for any $\psi_1 \in C_c(\mathbb{R}^d)$ with $\psi_1 \geq 0$, using again the uniqueness of the PDE in $\mathscr{L}_+$ we get

$$\int_{\mathbb{R}^d\times\Gamma_T} \varphi(e_{t_2}(\gamma)) \, d\nu_{x,t_1}(\gamma)\psi_1(x) \, d\mu_{t_1}^{\psi_0}(x) = \int_{\mathbb{R}^d\times\Gamma_T} \varphi(e_{t_2}(\gamma)) \, d\tilde{\nu}_{x,t_1}(\gamma)\psi_1(x) \, d\mu_{t_1}^{\psi_0}(x)$$

for any $\varphi \in C_c(\mathbb{R}^d)$, which can also be written as

$$\int_{\mathbb{R}^d\times\Gamma_T} \varphi(e_{t_2}(\gamma))\psi_1(e_{t_1}(\gamma)) \, d\nu_{x,t_1}(\gamma) \, d\mu_{t_1}^{\psi_0}(x) = \int_{\mathbb{R}^d\times\Gamma_T} \varphi(e_{t_2}(\gamma))\psi_1(e_{t_1}(\gamma)) \, d\tilde{\nu}_{x,t_1}(\gamma) \, d\mu_{t_1}^{\psi_0}(x). \tag{5.3.8}$$

Recalling that by (5.3.7)

$$\int_{\mathbb{R}^d} \nu_{x,t_1} \, d\mu_{t_1}^{\psi_0}(x) = \int_{\mathbb{R}^d} \nu_x \psi_0(x) \, d\mu_0(x), \quad \int_{\mathbb{R}^d} \tilde{\nu}_{x,t_1} \, d\mu_{t_1}^{\psi_0}(x) = \int_{\mathbb{R}^d} \tilde{\nu}_x \psi_0(x) \, d\mu_0(x),$$

by (5.3.8) we obtain

$$\int_{\mathbb{R}^d\times\Gamma_T} \varphi(e_{t_2}(\gamma))\psi_1(e_{t_1}(\gamma))\,d\nu_x(\gamma)\psi_0(x)\,d\mu_0(x)$$
$$= \int_{\mathbb{R}^d\times\Gamma_T} \varphi(e_{t_2}(\gamma))\psi_1(e_{t_1}(\gamma))\,d\tilde{\nu}_x(\gamma)\psi_0(x)\,d\mu_0(x)$$

for any non-negative $\psi_0, \psi_1, \varphi \in C_c(\mathbb{R}^d)$ (the constraint $\int_{\mathbb{R}^d}\psi_0\,d\mu_0 = 1$ can be easily removed multiplying the above equality by a positive constant). Iterating this argument, we finally get

$$\int_{\mathbb{R}^d\times\Gamma_T} \psi_n(e_{t_n}(\gamma))\ldots\psi_1(e_{t_1}(\gamma))\,d\nu_x(\gamma)\psi_0(x)\,d\mu_0(x)$$
$$= \int_{\mathbb{R}^d\times\Gamma_T} \psi_n(e_{t_n}(\gamma))\ldots\psi_1(e_{t_1}(\gamma))\,d\tilde{\nu}_x(\gamma)\psi_0(x)\,d\mu_0(x),$$

for any non-negative $\psi_0,\ldots,\psi_n \in C_c(\mathbb{R}^d)$, and thus (5.3.6) follows.
Considering now only rational times, we get that there exists a subset $A \subset \mathbb{R}^d$, with $\mu_0(A^c) = 0$, such that, for any $x \in A$,

$$(e_{t_1},\ldots,e_{t_n})_\#\nu_x = (e_{t_1},\ldots,e_{t_n})_\#\tilde{\nu}_x \quad \text{for any } t_1,\ldots,t_n \in [0,T]\cap\mathbb{Q}.$$

By continuity, this implies that, for any $x \in A$, $\nu_x = \tilde{\nu}_x$, as wanted. □

Remark 5.3.5. Suppose that forward uniqueness for the PDE holds in the class $\mathscr{L}_+$, and take $\mu_0 = \rho_0\mathscr{L}^d$ and $\tilde{\mu}_0 = \tilde{\rho}_0\mathscr{L}^d$, with $\rho_0, \tilde{\rho}_0 \in L^1_+(\mathbb{R}^d)\cap L^\infty_+(\mathbb{R}^d)$. If $\{\nu_x\}$ is a μ_0-SLF and $\{\tilde{\nu}_x\}$ is a $\tilde{\mu}_0$-SLF, then

$$\nu_x = \tilde{\nu}_x \quad \text{for } \mu_0 \wedge \tilde{\mu}_0\text{-a.e. } x.$$

In fact, by Remark 5.3.2 $\{\nu_x\}$ and $\{\tilde{\nu}_x\}$ are both $\mu_0\wedge\tilde{\mu}_0$-SLF, and thus we conclude by the uniqueness result proved above.

By Theorems 5.3.3 and 5.3.4, and by the remark above, we obtain the following:

Corollary 5.3.6 (Existence and uniqueness of SLF). *Let us assume that we have forward existence and uniqueness for the PDE in $\mathscr{L}_+$. Then there exists a measurable selection of martingale solution $\{\nu_x\}_{x\in\mathbb{R}^d}$ which is a μ_0-SLF for any $\mu_0 = \rho_0\mathscr{L}^d$ with $\rho_0 \in L^1_+(\mathbb{R}^d)\cap L^\infty_+(\mathbb{R}^d)$, and if $\{\tilde{\nu}_x\}_{x\in\mathbb{R}^d}$ is a $\tilde{\mu}_0$-SLF for a fixed $\tilde{\mu}_0 = \tilde{\rho}_0\mathscr{L}^d$ with $\tilde{\rho}_0 \in L^1_+(\mathbb{R}^d)\cap L^\infty_+(\mathbb{R}^d)$, then $\nu_x = \tilde{\nu}_x$ for $\mathscr{L}^d$-a.e. $x \in \operatorname{supp}(\tilde{\mu}_0)$.*

Proof. It suffices to consider a SLF starting from a Gaussian measure (which exists by Theorem 5.3.3), and to apply Remark 5.3.5. □

By now, the above selection of martingale solutions $\{\nu_x\}$, which is uniquely determined $\mathscr{L}^d$-a.e., will be called the SLF (starting at time 0 and relative to (b,a)).

We finally prove a stability result for SLF:

Theorem 5.3.7 (Stability of SLF starting from a fixed measure). *Let us suppose that $b^n, b : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $a^n, a : [0,T]\times\mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ are uniformly bounded functions, and that we have forward existence and uniqueness for the PDE in $\mathscr{L}_+$ with coefficients (b,a). Let $\mu_0 = \rho_0\mathscr{L}^d \in \mathcal{M}_+(\mathbb{R}^d)$, with $\rho_0 \in L^\infty(\mathbb{R}^d)$, and let $\{\nu^n_x\}_{x\in\mathbb{R}^d}$ and $\{\nu_x\}_{x\in\mathbb{R}^d}$ be μ_0-SLF for (b^n,a^n) and (b,a) respectively. Define $\nu^n := \int_{\mathbb{R}^d}\nu^n_x\,d\mu_0(x)$, $\nu := \int_{\mathbb{R}^d}\nu_x\,d\mu_0(x)$. Assume that:*

1. *$(b^n,a^n)\to(b,a)$ in $L^1_{loc}([0,T]\times\mathbb{R}^d)$;*
2. *setting $\rho^n_t\mathscr{L}^d = \mu^n_t := (e_t)_\#\nu^n$, for any $t\in[0,T]$*

$$\|\rho^n_t\|_{L^\infty(\mathbb{R}^d)} \le C \quad \text{for a certain constant } C = C(T).$$

Then $\nu^n \rightharpoonup^* \nu$ in $\mathcal{M}(\Gamma_T)$.

Proof. Since (b^n,a^n) are uniformly bounded in L^∞, as in Step 2 of the the proof of Theorem 5.2.7 one proves that the sequence of probability measures (ν^n) on $\mathbb{R}^d\times\Gamma_T$ is tight. In order to conclude, we must show that any limit point of (ν^n) is ν.

Let $\tilde\nu$ be any limit point of (ν^n). We claim that $\tilde\nu$ is concentrated on martingale solutions of the SDE with coefficients (b,a). Indeed, let us define $\tilde\mu_t := (e_t)_\#\tilde\nu$. Since $\mu^n_t\to\tilde\mu_t$ narrowly and ρ^n_t are non-negative functions bounded in $L^\infty(\mathbb{R}^d)$, we get $\tilde\mu_t = \rho_t\mathscr{L}^d$ for a certain non-negative function $\rho_t\in L^\infty(\mathbb{R}^d)$. We now observe that the argument used in Step 3 of the proof of Theorem 5.2.7 was using only the property that, for any $\varphi\in C^\infty_c(\mathbb{R}^d)$,

$$\limsup_{n\to+\infty}\sum_i\int_s^t\int_{\mathbb{R}^d}\Big|\Big(b^n_i(u,x)-\tilde b_i(u,x)\Big)\partial_i\varphi(x)\Big|\rho^n_u(x)\,dx\,du \le \sum_i\int_s^t\int_{\mathbb{R}^d}\Big|\Big(b_i(u,x)-\tilde b_i(u,x)\Big)\partial_i\varphi(x)\Big|\rho_u(x)\,dx\,du,$$

$$\limsup_{n\to+\infty}\sum_{ij}\int_s^t\int_{\mathbb{R}^d}\Big|\Big(a^n_{ij}(u,x)-\tilde a_{ij}(u,x)\Big)\partial_{ij}\varphi(x)\Big|\rho^n_u(x)\,dx\,du \le \sum_{ij}\int_s^t\int_{\mathbb{R}^d}\Big|\Big(a_{ij}(u,x)-\tilde a_{ij}(u,x)\Big)\partial_{ij}\varphi(x)\Big|\,\rho_u(x)\,dx\,du$$

for any $\tilde{b} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\tilde{a} : [0,T] \times \mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ bounded and continuous. This property simply follows by (i) and the w^*-convergence of ρ_t^n to ρ_t in $L^\infty([0,T] \times \mathbb{R}^d)$. Since $t \mapsto \rho_t \mathscr{L}^d$ is w^*-continuous in the sense of measures, the w^*-continuity of $t \mapsto \rho_t$ in $L^\infty(\mathbb{R}^d)$ follows. Thus, if we write $\tilde{\nu} := \int_{\mathbb{R}^d} \tilde{\nu}_x \, d\mu_0(x)$ (considering the disintegration of $\tilde{\nu}$ with respect to $\mu_0 = (e_0)_\# \tilde{\nu}$), we have proved that $\{\tilde{\nu}_x\}$ is a μ_0-SLF for (b,a). Therefore, by Theorem 5.3.4, we conclude that $\nu = \tilde{\nu}$. □

We remark that the theory just developed could be generalized to more general situations. Indeed the key property of the convex class $\mathscr{L}_+$ is the following monotonicy property:

$$0 \le \tilde{\mu}_t \le \mu_t \in \mathscr{L}_+ \quad \Rightarrow \quad \tilde{\mu}_t \in \mathscr{L}_+$$

(see also [5, Section 3]).

5.3.2 SLF versus RLF

We remark that, in the special case $a = 0$, our SLF coincides with a sort of superposition of the RLF introduced in [4]:

Lemma 5.3.8. *Let us assume $a = 0$. Then $\nu_{x,s}$ is a martingale solution of the SDE (which, in this case, is just an ODE) starting from x at time s if and only if it is concentrated on integral curves of the ODE, that is, for $\nu_{x,s}$-a.e. γ,*

$$\gamma(t) - \gamma(s) = \int_s^t b(\tau, \gamma(\tau)) \, d\tau \quad \forall t \in [s,T].$$

Proof. It is clear from the definition of martingale solution that, if $\nu_{x,s}$ is concentrated on integral curves on the ODE, then it is a martingale solution. Let us prove the converse implication. By the definition of martingale solution and the fact that $a = 0$, it is a known fact that

$$M_t := \gamma(t) - \gamma(s) - \int_s^t b(\tau, \gamma(\tau)) \, d\tau, \quad t \in [s,T],$$

is a $\nu_{x,s}$-martingale with zero quadratic variation. This implies that also M_t^2 is a martingale, and since $M_s = 0$ we get

$$0 = \mathbb{E}^{\nu_{x,s}}[M_t^2] = \int_{\Gamma_T} \left(\gamma(t) - \gamma(s) - \int_s^t b(\tau, \gamma(\tau)) \, d\tau \right)^2 d\nu_{x,s}(\gamma) \quad \forall t \in [s,T],$$

which gives the thesis. □

Thus, in the case $a = 0$, a martingale solution of the SDE starting from x is simply a measure on Γ_T concentrated on integral curves of b. By the results in [4] we know that, if we have forward uniqueness for the PDE in $\mathscr{L}_+$, then any measure ν on Γ_T concentrated on integral curves of b such that its time marginals induces a solution of the PDE in $\mathscr{L}_+$ is concentrated on a graph, i.e. there exists a function $x \mapsto X(\cdot, x) \in \Gamma_T$ such that

$$\nu = X(\cdot, x)_\# \mu_0, \quad \text{with} \quad \mu_0 := (e_0)_\# \nu$$

(see for instance [7, Theorem 18]). Then, if we assume forward uniqueness for the PDE in $\mathscr{L}_+$, our SLF coincides exactly with the RLF in [4]. Applying the stability result proved in the above paragraph, we obtain that, as the noise tends to 0, our SLF converges to the RLF associated to the ODE $\dot\gamma = b(\gamma)$. So we have a vanishing viscosity result for RLF.

Corollary 5.3.9. *Let us suppose that $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is uniformly bounded, and that we have forward existence and uniqueness for the PDE in $\mathscr{L}_+$ with coefficients $(b, 0)$. Let $\{\nu_x^\varepsilon\}_{x\in\mathbb{R}^d}$ and $\{\nu_x\}_{x\in\mathbb{R}^d}$ be the SLF relative to $(b, \varepsilon I)$ and $(b, 0)$ respectively (existence and uniqueness of martingale solutions for the SDE with coefficients $(b, \varepsilon I)$, together with the measurability of the family $\{\nu_x^\varepsilon\}_{x\in\mathbb{R}^d}$, follows by [125, Theorem 7.2.1]). Let $\mu_0 = \rho_0 \mathscr{L}^d \in \mathcal{M}_+(\mathbb{R}^d)$, with $\rho_0 \in L^\infty(\mathbb{R}^d)$, and define $\nu^\varepsilon := \int_{\mathbb{R}^d} \nu_x^\varepsilon \, d\mu_0(x)$, $\nu := \int_{\mathbb{R}^d} \nu_x \, d\mu_0(x)$.*
Set $\rho_t^\varepsilon \mathscr{L}^d = \mu_t^\varepsilon := (e_t)_\# \nu^\varepsilon$, and assume that for any $t \in [0,T]$

$$\|\rho_t^\varepsilon\|_{L^\infty(\mathbb{R}^d)} \le C \quad \text{for a certain constant } C = C(T).$$

Then $\nu^\varepsilon \rightharpoonup^* \nu$ in $\mathcal{M}(\Gamma_T)$.

In [4], the uniqueness of RLF implies the semigroup law (see [4], [5] for more details). In our case, by the uniqueness of SLF, we have as a consequence that the Chapman-Kolmogorov equation holds:

Proposition 5.3.10. *For any $s \ge 0$, let $\{\nu_{x,s}\}_{x\in\mathbb{R}^d}$ denotes the unique SLF starting at time s. Let us denote by $\nu_{s,x}(t, dy)$ the probability measure on $\mathbb{R}^d$ given by $\nu_{s,x}(t, \cdot) := (e_t)_\# \nu_{s,x}$. Then, for any $0 \le s < t < u \le T$,*

$$\int_{\mathbb{R}^d} \nu_{t,y}(u, \cdot) \nu_{s,x}(t, dy) = \nu_{s,x}(u, \cdot) \quad \text{for } \mathscr{L}^d - a.e. \ x.$$

Proof. Let us define

$$\tilde\nu_{s,x} := \begin{cases} \nu_{s,x} & \text{on } C([s,t], \mathbb{R}^d) \\ \int_{\mathbb{R}^d} \nu_{t,y} \nu_{s,x}(t, dy) & \text{on } C([t,T], \mathbb{R}^d). \end{cases}$$

This gives a family of martingale solution starting from x at time s (see [125]), and, using that $\{\nu_{x,s}\}$ and $\{\nu_{x,t}\}$ are SLF starting at time s and t respectively, it is simple to check that $\{\tilde\nu_{s,x}\}_{x\in\mathbb{R}^d}$ is a SLF starting at time s. Thus, by Theorem 5.3.4, we have the thesis. □

5.4 Fokker-Planck equation

We now want to study the Fokker-Planck equation

$$\partial_t \mu_t + \sum_i \partial_i (b_i \mu_t) - \frac{1}{2} \sum_{ij} \partial_{ij} (a_{ij} \mu_t) = 0 \quad \text{in } [0,T] \times \mathbb{R}^d, \tag{5.4.1}$$

where $a = (a_{ij})$ is symmetric and non-negative definite (that is, $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$).

5.4.1 Existence and uniqueness of measure valued solutions

Proposition 5.4.1. *Let us assume that $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ and $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ are bounded functions, having two bounded continuous spatial derivatives. Then, for any finite measure μ_0 there exists a unique finite measure-valued solution of (5.4.1) starting from μ_0 such that $|\mu_t|(\mathbb{R}^d) \le C$ for any $t \in [0,T]$.*

Proof. **Existence:** let $\{\nu_x\}_{x \in \mathbb{R}^d}$ be the measurable family of martingale solutions of the SDE

$$\begin{cases} dX = b(t,X)\,dt + \sqrt{a(t,X)}\,dB(t) \\ X(0) = x \end{cases}$$

(which exists and is unique by [125, Corollary 6.3.3]). Then, by Lemma 5.2.4 and Remark 5.2.6, the measure $\mu_t := (e_t)_\# \int_{\mathbb{R}^d} \nu_x \, d\mu_0(x)$ solves (5.4.1) and $|\mu_t|(\mathbb{R}^d) \le |\mu_0|(\mathbb{R}^d)$.
Uniqueness: by linearity, it suffices to prove that, if $\mu_0 = 0$, then $\mu_t = 0$ for all $t \in [0,T]$. Fix $\psi \in C_c^\infty(\mathbb{R}^d)$, $\bar t \in [0,T]$, and let $f(t,x)$ be the (unique) solution of

$$\begin{cases} \partial_t f + \sum_i b_i \partial_i f + \frac{1}{2} \sum_{ij} a_{ij} \partial_{ij} f = 0 & \text{in } [0,\bar t] \times \mathbb{R}^d \\ f(\bar t) = \psi & \text{on } \mathbb{R}^d \end{cases}$$

(which exists and is unique by [125, Theorem 3.2.6]). By [125, Theorems 3.1.1 and 3.2.4], we know that $f \in C_b^{1,2}$, i.e. it is uniformly bounded with one bounded continuous time derivative and two bounded continuous spatial derivatives. Since μ_t is a finite measure by assumption, and $t \mapsto \mu_t$ is narrowly continuous (Lemma 5.2.1), we can use $f(t,\cdot)$ as test functions in (5.1.3), and we get

$$\frac{d}{dt} \int_{\mathbb{R}^d} f(t,x)\,d\mu_t(x) = \int_{\mathbb{R}^d} \Big[\partial_t f(t,x) + \sum_i b_i(t,x) \partial_i f(t,x) + \frac{1}{2} \sum_{ij} a_{ij}(t,x) \partial_{ij} f(t,x) \Big] \, d\mu_t(x) = 0$$

(the above computation is admissible since $f \in C_b^{1,2}$). This implies in particular that

$$0 = \int_{\mathbb{R}^d} f(0,x)\,d\mu_0(x) = \int_{\mathbb{R}^d} f(\bar t, x)\,d\mu_{\bar t}(x) = \int_{\mathbb{R}^d} \psi(x)\,d\mu_{\bar t}(x).$$

By the arbitrariness of ψ and $\bar t$ we obtain $\mu_t = 0$ for all $t \in [0,T]$. ∎

We remark that, in the uniformly parabolic case, the above proof still works under weaker regularity assumptions. Indeed, in that case, one has existence of a measurable family of martingale solutions of the SDE and of a solution $f \in C_b^{1,2}([0,\bar{t}] \times \mathbb{R}^d)$ of the adjoint equation if a and b are just Hölder continuous (see [125, Theorem 3.2.1]). So we get:

Proposition 5.4.2. *Let us assume that* $a : [0,T]\times\mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ *and* $b : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ *are bounded functions such that:*

1. $\langle \xi, a(t,x)\xi\rangle \geq \alpha|\xi|^2\ \forall (t,x) \in [0,T]\times\mathbb{R}^d$, *for some* $\alpha > 0$;
2. $|b(t,x)-b(s,y)|+\|a(t,x)-a(s,y)\| \leq C\left(|x-y|^\delta + |t-s|^\delta\right) \forall (t,x),(s,y) \in [0,T]\times\mathbb{R}^d$, *for some* $\delta \in (0,1]$, $C \geq 0$.

Then, for any finite measure μ_0 *there exists a unique finite measure-valued solution of (5.4.1) starting from* μ_0.

5.4.2 Existence and uniqueness of absolutely continuous solutions in the uniformly parabolic case

We are now interested in absolutely continuous solutions of (5.1.2). Therefore, we consider the following equation

$$\begin{cases} \partial_t u + \sum_i \partial_i(b_i u) - \frac{1}{2}\sum_{ij}\partial_{ij}(a_{ij}u) = 0 \quad \text{in } [0,T]\times\mathbb{R}^d, \\ u(0) = u_0, \end{cases} \tag{5.4.2}$$

which must be understood in the distributional sense on $[0,T]\times\mathbb{R}^d$. We now first prove an existence and uniqueness result in the L^2-setting under a regularity assumption on the divergence of a, which enables us to write (5.4.2) in a variational form, and thus to apply classical existence results (the uniqueness part in L^2 is much more involved). After, we will give a maximum principle result.
Let us make the following assumptions on the coefficients:

$$\begin{aligned} &\sum_j \partial_j a_{ij} \in L^\infty([0,T]\times\mathbb{R}^d) \ \text{ for } i=1,\dots,d, \quad (\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij})^- \in L^\infty([0,T]\times\mathbb{R}^d), \\ &\langle \xi, a(t,x)\xi\rangle \geq \alpha|\xi|^2 \ \ \forall (t,x) \in [0,T]\times\mathbb{R}^d, \quad \text{for some } \alpha > 0. \end{aligned} \tag{5.4.3}$$

Theorem 5.4.3. *Let us assume that* $a : [0,T]\times\mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ *and* $b : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ *are bounded functions such that (5.4.3) is fulfilled. Then, for any* $u_0 \in L^2(\mathbb{R}^d)$, *(5.4.2) has a unique solution* $u \in Y$, *where*

$$Y := \left\{u \in L^2([0,T],H^1(\mathbb{R}^d)) \mid \partial_t u \in L^2([0,T],H^{-1}(\mathbb{R}^d))\right\}.$$

If moreover $\partial_t a_{ij} \in L^\infty([0,T] \times \mathbb{R}^d)$ for $i,j = 1,\dots,d$, then existence and uniqueness holds in $L^2([0,T] \times \mathbb{R}^d)$, and so in particular any solution $u \in L^2([0,T] \times \mathbb{R}^d)$ of (5.4.2) belongs to Y.

The proof the above theorem is quite standard, except for the uniqueness result in the large space L^2, which is indeed quite technical and involved. The motivation for this more general result is that $L^1_+(\mathbb{R}^d) \cap L^\infty_+(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$, and $L^1_+(\mathbb{R}^d) \cap L^\infty_+(\mathbb{R}^d)$ is the space where we need well-posedness of the PDE if we want to apply the theory on martingale solutions developed in the last section (see Theorems 5.1.3 and 5.5.1).

We now give some properties of the family of solutions of (5.4.2):

Proposition 5.4.4. *We assume that $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ and $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ are bounded functions, and that (5.4.3) is fulfilled. Then the solution $u \in Y$ provided by Theorem 5.4.3 satisfies:*

(a) $u_0 \geq 0 \quad \Rightarrow \quad u \geq 0$;

(b) $u_0 \in L^\infty(\mathbb{R}^d) \quad \Rightarrow \quad u \in L^\infty([0,T] \times \mathbb{R}^d)$ *and we have*

$$\|u(t)\|_{L^\infty(\mathbb{R}^d)} \leq \|u_0\|_{L^\infty(\mathbb{R}^d)} e^{t\|(\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij})^-\|_\infty};$$

(c) if moreover

$$\frac{a}{1+|x|^2} \in L^2([0,T] \times \mathbb{R}^d), \qquad \frac{b}{1+|x|} \in L^2([0,T] \times \mathbb{R}^d),$$

then $u_0 \in L^1 \quad \Rightarrow \quad \|u(t)\|_{L^1(\mathbb{R}^d)} \leq \|u_0\|_{L^1(\mathbb{R}^d)} \quad \forall t \in [0,T]$.

We observe that, by the above results together with Proposition 5.4.2, we obtain:

Corollary 5.4.5. *Let us assume that $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}(\mathbb{R}^d)$ and $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ are bounded functions such that:*

1. $\langle \xi, a(t,x)\xi \rangle \geq \alpha |\xi|^2 \; \forall (t,x) \in [0,T] \times \mathbb{R}^d$, *for some* $\alpha > 0$;

2. $|b(t,x) - b(s,y)| + \|a(t,x) - a(s,y)\| \leq C\left(|x-y|^\gamma + |t-s|^\gamma\right) \forall (t,x),(s,y) \in [0,T] \times \mathbb{R}^d$, *for some* $\gamma \in (0,1]$, $C \geq 0$;

3. $\sum_j \partial_j a_{ij} \in L^\infty([0,T] \times \mathbb{R}^d)$ *for* $i = 1,\dots,d$, $\quad (\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij})^- \in L^\infty([0,T] \times \mathbb{R}^d)$;

4. $\frac{a}{1+|x|^2} \in L^2([0,T] \times \mathbb{R}^d)$, $\frac{b}{1+|x|} \in L^2([0,T] \times \mathbb{R}^d)$.

Then, for any $\mu_0 \in \mathcal{M}_+(\mathbb{R}^d)$ there exists a unique finite measure-valued solution $\mu_t \in \mathcal{M}_+(\mathbb{R}^d)$ of (5.1.2) starting from μ_0. Moreover, if such that $\mu_0 = \rho_0 \mathscr{L}^d$ with $\rho_0 \in L^2(\mathbb{R}^d)$, then $\mu_t \ll \mathscr{L}^d$ for all $t \in [0,T]$.

Proof. Existence and uniqueness of finite measure-valued solutions follows by Proposition 5.4.2. So the only thing to prove is that, if $\rho_0 \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ is non-negative, then $\mu_t \in \mathcal{M}_+(\mathbb{R}^d)$ and $\mu_t \ll \mathscr{L}^d$ for all $t \in [0,T]$. This simply follows by the fact that the solution $u \in Y$ provided by Theorem 5.4.3 belongs to $L^1_+(\mathbb{R}^d)$ by Proposition 5.4.4, and thus coincides with μ_t by uniqueness in the set of finite measure-valued solutions. □

In order to prove the results stated before, we need the following theorem of J.-L.Lions (see [93]):

Theorem 5.4.6. *Let H be an Hilbert space, provided with a norm $|\cdot|$, and inner product $(\cdot,\cdot)$. Let $\Phi \subset H$ be a subspace endowed with a prehilbertian norm $\|\cdot\|$, such that the injection $\Phi \hookrightarrow H$ is continuous. We consider a bilinear form $B : H \times \Phi \to \mathbb{R}$ such that:*

- *$H \ni u \mapsto B(u,\varphi)$ is continuous on H for any fixed $\varphi \in \Phi$;*
- *there exists $\alpha > 0$ such that $B(\varphi,\varphi) \geq \alpha \|\varphi\|^2$ for any $\varphi \in \Phi$.*

Then, for any linear continuous form L on Φ there exists $v \in H$ such that

$$B(v,\varphi) = L(\varphi) \quad \forall \varphi \in \Phi.$$

Proof of Theorem 5.4.3. We will first prove existence and uniqueness of a solution in the space Y. Once this will be done, we will show that, if u is a weak solution of (5.4.2) belonging to $L^2([0,T]\times\mathbb{R}^d)$ and $\partial_t a_{ij} \in L^\infty([0,T]\times\mathbb{R}^d)$ for $i,j = 1,\dots,d$, then u belongs to Y, and so it coincides with the unique solution provided before.
The change of unknown

$$v(t,x) = e^{-\lambda t} u(t,x)$$

leads to the equation

$$\begin{cases} \partial_t v + \sum_i \partial_i(\tilde{b}_i v) - \frac{1}{2}\sum_{ij} \partial_i(a_{ij}\partial_j v) + \lambda v = 0 & \text{in } [0,T]\times\mathbb{R}^d, \\ v_0 = u_0, \end{cases} \tag{5.4.4}$$

where $\tilde{b}_i := b_i - \frac{1}{2}\sum_j \partial_j a_{ij} \in L^\infty([0,T]\times\mathbb{R}^d)$. Assuming that λ satisfies $\lambda > \frac{1}{2}\|(\sum_i \partial_i \tilde{b}_i)^-\|_\infty$, we will prove existence and uniqueness for u.
Step 1: existence in Y. We want to apply Theorem 5.4.6.
Let us take $H := L^2([0,T], H^1(\mathbb{R}^d))$, $\Phi := \{\varphi \in C^\infty([0,T]\times\mathbb{R}^d) \mid \operatorname{supp}\varphi \subset\subset [0,T)\times\mathbb{R}^d\}$.
Φ is endowed with the norm

$$\|\varphi\|_\Phi^2 := \|\varphi\|_H^2 + \frac{1}{2}\int_{\mathbb{R}^d} |\varphi(0,x)|^2\,dx.$$

The bilinear form B and the linear form L are defined as

$$B(u,\varphi) := \int_0^T \int_{\mathbb{R}^d} \Big[u\Big(-\partial_t\varphi - \sum_i \tilde{b}_i\partial_i\varphi + \lambda\varphi\Big) + \frac{1}{2}\sum_{ij} a_{ij}\partial_j u \partial_i\varphi\Big]\, dx\, dt,$$

$$L(\varphi) := \int_{\mathbb{R}^d} u_0(x)\varphi(0,x)\, dx.$$

Thanks to these definitions and our assumptions, Lions' theorem applies, and we find a distributional solution v of (5.4.4). In particular,

$$\partial_t v = -\sum_i \partial_i(\tilde{b}_i v) + \frac{1}{2}\sum_{ij}\partial_i(a_{ij}\partial_j v) - \lambda v \in H^* = L^2([0,T], H^{-1}(\mathbb{R}^d)),$$

and thus $v \in Y$. In order to give a meaning to the initial condition and to show the uniqueness, we recall that for functions in Y there exists a well-defined notion of trace at 0 in $L^2(\mathbb{R}^d)$, and the following Gauss-Green formula holds:

$$\int_0^T \int_{\mathbb{R}^d} \partial_t u\tilde{u} + \partial_t\tilde{u}u\, dx\, dt = \int_{\mathbb{R}^d} u(T,x)\tilde{u}(T,x)\, dx - \int_{\mathbb{R}^d} u(0,x)\tilde{u}(0,x)\, dx \quad \forall u,\tilde{u} \in Y \tag{5.4.5}$$

(both facts follow by a standard approximation with smooth functions and by the fact that, if u is smooth and compactly supported in $[0,T) \times \mathbb{R}^d$, $\int_{\mathbb{R}^d} u^2(0,x)\, dx \le 2\|\partial_t u\|_{H^*}\|u\|_H$). Thus, by (5.4.4) and (5.4.5), we obtain that v satisfies

$$\int_{\mathbb{R}^d} (v(0,x) - u_0(x))\varphi(0,x)\, dx = 0 \quad \forall\varphi \in \Phi,$$

and therefore the initial condition is satisfied in $L^2(\mathbb{R}^d)$.

Step 2: uniqueness in Y. For the uniqueness, if $v \in Y$ is a solution of (5.4.4) with $u_0 = 0$, again by (5.4.5) we get

$$\begin{aligned}
0 &= \int_0^T \int_{\mathbb{R}^d} \Big(\partial_t v + \sum_i \partial_i(\tilde{b}_i v) - \frac{1}{2}\sum_{ij}\partial_i(a_{ij}\partial_j v) + \lambda v\Big) v\, dx\, dt \\
&= \frac{1}{2}\int_0^T \int_{\mathbb{R}^d} \Big[\frac{d}{dt}v^2 - \sum_i \tilde{b}_i\partial_i(v^2) + \sum_{ij} a_{ij}\partial_i v\partial_j v + 2\lambda v^2\Big]\, dx \\
&\ge \frac{1}{2}\int_{\mathbb{R}^d} v^2(T,x)\, dx + \Big(\lambda - \frac{1}{2}\|(\sum_i \partial_i\tilde{b}_i)^-\|_\infty\Big)\int_0^T \int_{\mathbb{R}^d} v^2\, dx\, dt \\
&\ge \Big(\lambda - \frac{1}{2}\|(\sum_i \partial_i\tilde{b}_i)^-\|_\infty\Big)\int_0^T \int_{\mathbb{R}^d} v^2\, dx\, dt.
\end{aligned}$$

Since $\lambda > \frac{1}{2}\|(\sum_i \partial_i\tilde{b}_i)^-\|_\infty$, we get $v = 0$.

Remark 5.4.7. We observe that the above proof still works for the PDE

$$\begin{cases} \partial_t u + \sum_i \partial_i(b_i u) - \frac{1}{2}\sum_{ij}\partial_{ij}(a_{ij}u) = U & \text{in } [0,T]\times\mathbb{R}^d, \\ u(0) = u_0, \end{cases}$$

with $U \in H^* = L^2([0,T], H^{-1}(\mathbb{R}^d))$. Indeed, it suffices to define L as

$$L(\varphi) := \langle U, \varphi\rangle_{H^*,H} + \int_{\mathbb{R}^d} u_0(x)\varphi(x)\,dx,$$

and all the rest of the proof works without any changes.

Thanks to this remark, we can now prove uniqueness in the larger space $L^2([0,T]\times\mathbb{R}^d)$ under the assumption $\partial_t a_{ij} \in L^\infty([0,T]\times\mathbb{R}^d)$ for $i,j = 1,\dots,d,$.
Step 3: uniqueness in L^2. If $u \in L^2([0,T]\times\mathbb{R}^d)$ is a (distributional) solution of (5.4.1), then

$$\partial_t u - \frac{1}{2}\sum_{ij}\partial_i(a_{ij}\partial_j u) = -\sum_i \partial_i(\tilde{b}_i u) \in L^2([0,T], H^{-1}(\mathbb{R}^d)).$$

By Remark 5.4.7, there exists $\tilde{u} \in Y$ solution of the above equation, with the same initial condition. Let us define $w := u - \tilde{u} \in L^2([0,T]\times\mathbb{R}^d)$. Then w is a distributional solution of

$$\begin{cases} \partial_t w - A(\partial_x)w := \partial_t w - \frac{1}{2}\sum_{ij}\partial_i(a_{ij}\partial_j w) = 0 & \text{in } [0,T]\times\mathbb{R}^d, \\ w(0) = 0. \end{cases}$$

In order to conclude the proof, it suffices to prove that $w = 0$.
Step 3.1: regularization. Let us consider the PDE

$$w_\varepsilon - \varepsilon A(\partial_x)w_\varepsilon = w \quad \text{in } [0,T]\times\mathbb{R}^d \tag{5.4.6}$$

(this is an elliptic problem degenerate in the time variable). Applying Theorem 5.4.6, with $H = \Phi := L^2([0,T], H^1(\mathbb{R}^d))$,

$$B(u,\varphi) := \int_0^T \int_{\mathbb{R}^d} \Big(u\varphi + \frac{\varepsilon}{2}\sum_{ij} a_{ij}\partial_j u \partial_i \varphi\Big)\,dx\,dt,$$

$$L(\varphi) := \int_0^T \int_{\mathbb{R}^d} w\varphi\,dx\,dt,$$

we find a unique solution w_ε of (5.4.6) in $L^2([0,T], H^1(\mathbb{R}^d))$, that is $w_\varepsilon = (I - \varepsilon A(\partial_x))^{-1}w$, with $(I - \varepsilon A(\partial_x)) : L^2([0,T], H^1(\mathbb{R}^d)) \to L^2([0,T], H^{-1}(\mathbb{R}^d))$ isomorphism. Now we want to find the equation solved by w_ε. We observe that, since $(I-\varepsilon A(\partial_x))^{-1}$ commutes with $A(\partial_x)$ and $\partial_t w = A(\partial_x)w$, the parabolic equation solved by w_ε formally looks

$$\partial_t w_\varepsilon - A(\partial_x)w_\varepsilon = [\partial_t, (I-\varepsilon A(\partial_x))^{-1}]w.$$

Formally computing the commutator between ∂_t and $(I-\varepsilon A(\partial_x))^{-1}$, one obtains

$$\partial_t w_\varepsilon - A(\partial_x)w_\varepsilon = \varepsilon(I-\varepsilon A(\partial_x))^{-1}\sum_{ij}\partial_j(\partial_t a_{ij}\partial_i w^\varepsilon) \tag{5.4.7}$$

in the distributional sense (see (5.4.9) below). Let us assume for a moment that (5.4.7) has been rigorously justified, and let us see how we can conclude.

Step 3.2: Gronwall argument. By (5.4.7) it follows that $\partial_t w_\varepsilon \in L^2([0,T], H^{-1}(\mathbb{R}^d))$. Thus, recalling that $w_\varepsilon \in L^2([0,T], H^1(\mathbb{R}^d))$, we can multiply (5.4.7) by w_ε and integrate on $\mathbb{R}^d$, obtaining

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^d}|w_\varepsilon|^2\,dx + \alpha\int_{\mathbb{R}^d}|\nabla_x w_\varepsilon|^2\,dx \le -\varepsilon\int_{\mathbb{R}^d}\sum_{ij}(\partial_t a_{ij})\partial_i w_\varepsilon \partial_j\big((I-\varepsilon A(\partial_x))^{-1}w_\varepsilon\big)\,dx.$$

We observe that $w_\varepsilon(t)\to 0$ in L^2 as $t\searrow 0$. Indeed, since $w_\varepsilon\in Y$ there is a well-defined notion of trace at 0 in L^2 (see (5.4.5)), and it is not difficult to see that this trace is 0 since $w(0)=0$ in the sense of distributions. Thus, integrating in time the above inequality, we get

$$\begin{aligned}&\|w_\varepsilon(t)\|^2_{L^2(\mathbb{R}^d)} + 2\alpha\|\nabla_x w_\varepsilon\|^2_{L^2([0,T]\times\mathbb{R}^d)}\\ &\quad\le 2C\varepsilon\|\nabla_x w_\varepsilon\|_{L^2([0,T]\times\mathbb{R}^d)}\|\nabla_x\big((I-\varepsilon A(\partial_x))^{-1}w_\varepsilon\big)\|_{L^2([0,T]\times\mathbb{R}^d)}\quad \forall t\in[0,T].\end{aligned} \tag{5.4.8}$$

Let us consider, for a general $v\in L^2$, the function $v_\varepsilon := (I-\varepsilon A(\partial_x))^{-1}v$. Multiplying the identity $v_\varepsilon - \varepsilon A(\partial_x)v_\varepsilon = v$ by v_ε and integrating on $[0,T]\times\mathbb{R}^d$, we get

$$\|v_\varepsilon\|^2_{L^2} + \alpha\varepsilon\|\nabla_x v_\varepsilon\|^2_{L^2} \le \|v_\varepsilon\|_{L^2}\|v\|_{L^2},$$

which implies $\|v_\varepsilon\|_{L^2}\le\|v\|_{L^2}$, and therefore $\alpha\varepsilon\|\nabla_x v_\varepsilon\|^2_{L^2}\le\|v\|^2_{L^2}$. Applying this last inequality with $v=w_\varepsilon$, we obtain

$$\|\nabla_x\big((I-\varepsilon A(\partial_x))^{-1}w_\varepsilon\big)\|_{L^2([0,T]\times\mathbb{R}^d)} \le \frac{1}{\sqrt{\alpha\varepsilon}}\|w_\varepsilon\|_{L^2([0,T]\times\mathbb{R}^d)}.$$

Substituting the above inequality in (5.4.8), we have

$$\begin{aligned}\|w_\varepsilon(t)\|^2_{L^2(\mathbb{R}^d)} + 2\alpha\|\nabla_x w_\varepsilon\|^2_{L^2([0,T]\times\mathbb{R}^d)} &\le 2C\sqrt{\frac{\varepsilon}{\alpha}}\|\nabla_x w_\varepsilon\|_{L^2([0,T]\times\mathbb{R}^d)}\|w_\varepsilon\|_{L^2([0,T]\times\mathbb{R}^d)}\\ &\le C\sqrt{\frac{\varepsilon}{\alpha}}\|\nabla_x w_\varepsilon\|^2_{L^2([0,T]\times\mathbb{R}^d)} + C\sqrt{\frac{\varepsilon}{\alpha}}\|w_\varepsilon\|^2_{L^2([0,T]\times\mathbb{R}^d)},\end{aligned}$$

which implies, for ε small enough (say $\varepsilon \le 4\frac{\alpha^3}{C^2}$),

$$\|w_\varepsilon(t)\|^2_{L^2(\mathbb{R}^d)} \le C\sqrt{\frac{\varepsilon}{\alpha}}\|w_\varepsilon\|^2_{L^2([0,T]\times\mathbb{R}^d)} \quad \forall t \in [0,T].$$

By Gronwall inequality $w_\varepsilon = 0$, and thus by (5.4.6) $w = 0$.

Step 3.3: rigorous justification of (5.4.7). In order to conclude the proof of the theorem, we only need to rigorously justify (5.4.7).

Let $(a^n_{ij})_{n\in\mathbb{N}}$ be a sequence of smooth functions bounded in L^∞, such that $\langle a^n\xi, \xi\rangle \ge \frac{\alpha}{2}|\xi|^2$, $\sum_j \partial_j a^n_{ij}$ and $\partial_t a^n_{ij}$ are uniformly bounded, and $a^n_{ij} \to a_{ij}$, $\sum_j \partial_j a^n_{ij} \to \sum_j \partial_j a_{ij}$, $\partial_t a^n_{ij} \to \partial_t a_{ij}$ a.e.

We now compute $[\partial_t, (I - \varepsilon A^n(\partial_x))^{-1}]$, where $A^n(\partial_x) := \sum_{ij} \partial_i(a^n_{ij}\partial_j\cdot)$:

$$\begin{aligned}
[\partial_t, (I - \varepsilon A^n(\partial_x))^{-1}] &= [\partial_t, \sum_{k\ge 0} \varepsilon^k A^n(\partial_x)^k] = \sum_{n\ge 0} \varepsilon^k [\partial_t, A^n(\partial_x)^k] \\
&= \varepsilon \sum_{k=0}^{\infty}\sum_{i=0}^{k-1} \big(\varepsilon A^n(\partial_x)\big)^i [\partial_t, A^n(\partial_x)] \big(\varepsilon A^n(\partial_x)\big)^{k-i-1} \\
&= \varepsilon \sum_{i=0}^{\infty} \big(\varepsilon A^n(\partial_x)\big)^i [\partial_t, A^n(\partial_x)] \sum_{k>i} \big(\varepsilon A^n(\partial_x)\big)^{k-i-1} \\
&= \varepsilon (I - \varepsilon A^n(\partial_x))^{-1} [\partial_t, A^n(\partial_x)] (I - \varepsilon A^n(\partial_x))^{-1},
\end{aligned} \tag{5.4.9}$$

where at the second equality we used the algebraic identity $[A, B^k] = \sum_{i=0}^{k-1} B^i[A,B]B^{k-i-1}$. Thus, for any $\varphi, \psi \in C^\infty_c([0,T]\times\mathbb{R}^d)$, we have

$$\begin{aligned}
\int_0^T\int_{\mathbb{R}^d} \psi\partial_t \left((I - \varepsilon A^n(\partial_x))^{-1}\varphi\right)\, dx\, dt &= \int_0^T\int_{\mathbb{R}^d} \psi \left[(I - \varepsilon A^n(\partial_x))^{-1}\partial_t\varphi\right]\, dx\, dt \\
+ \varepsilon \int_0^T\int_{\mathbb{R}^d} \psi &\left[(I - \varepsilon A^n(\partial_x))^{-1}[\partial_t, A^n(\partial_x)](I - \varepsilon A^n(\partial_x))^{-1}\varphi\right]\, dx\, dt.
\end{aligned} \tag{5.4.10}$$

We now want to pass to the limit in the above identity as $n \to \infty$. Since $(I - \varepsilon A^n(\partial_x))^{-1}$ is selfadjoint in $L^2([0,T]\times\mathbb{R}^d)$ and it commutes with $A^n(\partial_x)$, we get

$$\begin{aligned}
&\int_0^T\int_{\mathbb{R}^d} \psi \left[(I - \varepsilon A^n(\partial_x))^{-1}[\partial_t, A^n(\partial_x)](I - \varepsilon A^n(\partial_x))^{-1}\varphi\right]\, dx\, dt \\
&\quad = \int_0^T\int_{\mathbb{R}^d} \left[(I - \varepsilon A^n(\partial_x))^{-1}\psi\right] \left[[\partial_t, A^n(\partial_x)](I - \varepsilon A^n(\partial_x))^{-1}\varphi\right]\, dx\, dt \\
&\quad = -\int_0^T\int_{\mathbb{R}^d} \left[\partial_t\big((I - \varepsilon A^n(\partial_x))^{-1}\psi\big)\right] \left[(I - \varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi\right]\, dx\, dt \\
&\quad - \int_0^T\int_{\mathbb{R}^d} \left[(I - \varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\psi\right] \left[\partial_t\big((I - \varepsilon A^n(\partial_x)^{-1})\varphi\big)\right]\, dx\, dt.
\end{aligned}$$

By (5.4.9) we have

$$\partial_t\big((I-\varepsilon A^n(\partial_x))^{-1}\varphi\big) = (I-\varepsilon A^n(\partial_x))^{-1}\partial_t\varphi+\varepsilon(I-\varepsilon A^n(\partial_x))^{-1}[\partial_t, A^n(\partial_x)](I-\varepsilon A^n(\partial_x))^{-1}\varphi,$$

and, observing that $[\partial_t, A^n(\partial_x)] = \sum_{ij} \partial_i(\partial_t a^n_{ij}\partial_j\cdot)$, we deduce that the right hand side is uniformly bounded in $L^2([0,T], H^1(\mathbb{R}^d))$. In the same way one obtains

$$\begin{aligned}\partial_t\big((I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi\big) &= (I-\varepsilon A^n(\partial_x))^{-1}\partial_t\big(A^n(\partial_x)\varphi\big)\\ &\quad+\varepsilon(I-\varepsilon A^n(\partial_x))^{-1}[\partial_t, A^n(\partial_x)](I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi\\ &= (I-\varepsilon A^n(\partial_x))^{-1}[\partial_t, A^n(\partial_x)]\varphi\\ &\quad+(I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\partial_t\varphi\\ &\quad+\varepsilon(I-\varepsilon A^n(\partial_x))^{-1}[\partial_t, A^n(\partial_x)](I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi,\end{aligned}$$

and, as above, the right hand side is uniformly bounded in $L^2([0,T], H^1(\mathbb{R}^d))$. Thus $\partial_t(I-\varepsilon A^n(\partial_x))^{-1}\varphi$ is uniformly bounded in $L^2([0,T], H^1(\mathbb{R}^d)) \subset L^2([0,T]\times\mathbb{R}^d)$ (the same obviously holds for ψ in place of φ), while $(I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi$ is uniformly bounded in $H^1([0,T]\times\mathbb{R}^d)$ (again the same fact holds for ψ in place of φ). Therefore, since $H^1_{loc}([0,T]\times\mathbb{R}^d) \hookrightarrow L^2_{loc}([0,T]\times\mathbb{R}^d)$ compactly, all we have to check is that

$$\partial_t\big((I-\varepsilon A^n(\partial_x))^{-1}\varphi\big) \to \partial_t\big((I-\varepsilon A(\partial_x))^{-1}\varphi\big)$$

and

$$(I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi \to (I-\varepsilon A(\partial_x))^{-1}A(\partial_x)\varphi$$

in the sense of distribution (indeed, by what we have shown above, $\partial_t\big((I-\varepsilon A^n(\partial_x))^{-1}\varphi\big)$ will converge weakly in L^2 while $(I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi$ will converge strongly in L^2_{loc}, and therefore it is not difficult to see that the product converges to the product of the limits). We observe that, since the solution of

$$\varphi_\varepsilon - \varepsilon A(\partial_x)\varphi_\varepsilon = \varphi \quad \text{in } [0,T]\times\mathbb{R}^d \tag{5.4.11}$$

belonging to $L^2([0,T], H^1(\mathbb{R}^d))$ is unique, and any limit point of $(I-\varepsilon A^n(\partial_x))^{-1}\varphi$ belongs to $L^2([0,T], H^1(\mathbb{R}^d))$ and is a distributional solution of (5.4.11), one obtains that

$$(I-\varepsilon A^n(\partial_x))^{-1}\varphi \to (I-\varepsilon A(\partial_x))^{-1}\varphi$$

in the distributional sense, which implies the convergence of $\partial_t(I-\varepsilon A^n(\partial_x))^{-1}\varphi$ to $\partial_t(I-\varepsilon A(\partial_x))^{-1}\varphi$. Regarding $(I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi$, let us take $\chi \in C^\infty_c([0,T]\times\mathbb{R}^d)$. Then we consider

$$\int_0^T\int_{\mathbb{R}^d} A^n(\partial_x)\varphi\,\big[(I-\varepsilon A^n(\partial_x))^{-1}\chi\big]\,dx\,dt = -\int_0^T\int_{\mathbb{R}^d}\sum_{ij} a^n_{ij}\partial_j\varphi\,\big(\partial_i(I-\varepsilon A^n(\partial_x))^{-1}\chi\big)\,dx\,dt.$$

Recalling that $(I-\varepsilon A^n(\partial_x))^{-1}\chi$ is uniformly bounded in $L^2([0,T],H^1(\mathbb{R}^d))$, we get that $\partial_j(I-\varepsilon A^n(\partial_x))^{-1}\chi$ converges to $\partial_j(I-\varepsilon A(\partial_x))^{-1}\chi$ weakly in $L^2([0,T]\times\mathbb{R}^d)$ while $a^n_{ij}\to a_{ij}$ a.e., and so the convergence of $(I-\varepsilon A^n(\partial_x))^{-1}A^n(\partial_x)\varphi$ to $(I-\varepsilon A(\partial_x))^{-1}A(\partial_x)\varphi$ follows.
Thus we are able to pass to the limit in (5.4.10), and we get $\partial_t\big((I-\varepsilon A(\partial_x))^{-1}\varphi\big)\in L^2([0,T],H^1(\mathbb{R}^d))$ and

$$\begin{aligned}\int_0^T\int_{\mathbb{R}^d}\psi\partial_t\left((I-\varepsilon A(\partial_x))^{-1}\varphi\right)\,dx\,dt &= \int_0^T\int_{\mathbb{R}^d}\psi\left[(I-\varepsilon A(\partial_x))^{-1}\partial_t\varphi\right]\,dx\,dt\\ &+\varepsilon\int_0^T\int_{\mathbb{R}^d}\psi\left[(I-\varepsilon A(\partial_x))^{-1}[\partial_t,A(\partial_x)](I-\varepsilon A(\partial_x))^{-1}\varphi\right]\,dx\,dt.\end{aligned}$$

Observing that $(I-\varepsilon A(\partial_x))^{-1}$ is selfadjoint in $L^2([0,T]\times\mathbb{R}^d)$ (for instance, this can be easily proved by approximation), we have that the second integral in the right hand side can be written as

$$\begin{aligned}&\int_0^T\int_{\mathbb{R}^d}\psi\left[(I-\varepsilon A(\partial_x))^{-1}[\partial_t,A(\partial_x)](I-\varepsilon A(\partial_x))^{-1}\varphi\right]\,dx\,dt\\ &\qquad=\int_0^T\int_{\mathbb{R}^d}\left[(I-\varepsilon A(\partial_x))^{-1}\psi\right]\left[[\partial_t,A(\partial_x)]\left((I-\varepsilon A(\partial_x))^{-1}\varphi\right)\right]\,dx\,dt.\end{aligned}$$

Using now that $[\partial_t,A(\partial_x)]=\sum_{ij}\partial_i(\partial_t a_{ij}\partial_j\cdot)$ in the sense of distributions, it can be easily proved by approximation that the right hand side above coincides with

$$-\int_0^T\int_{\mathbb{R}^d}\sum_{ij}(\partial_t a_{ij})\partial_i\left((I-\varepsilon A(\partial_x))^{-1}\psi\right)\partial_j\left((I-\varepsilon A(\partial_x))^{-1}\varphi\right)\,dx\,dt.$$

Therefore we finally obtain

$$\begin{aligned}\int_0^T\int_{\mathbb{R}^d}\psi\partial_t\left((I-\varepsilon A(\partial_x))^{-1}\varphi\right)\,dx\,dt &= \int_0^T\int_{\mathbb{R}^d}\psi\left[(I-\varepsilon A(\partial_x))^{-1}\partial_t\varphi\right]\,dx\,dt\\ -\varepsilon\int_0^T\int_{\mathbb{R}^d}&\sum_{ij}(\partial_t a_{ij})\partial_i\left((I-\varepsilon A(\partial_x))^{-1}\psi\right)\partial_j\left((I-\varepsilon A(\partial_x))^{-1}\varphi\right)\,dx\,dt.\end{aligned}\tag{5.4.12}$$

By what we have proved above, it follows that

$$\begin{aligned}&\partial_t\big((I-\varepsilon A(\partial_x))^{-1}\varphi\big)\in L^2([0,T],H^1(\mathbb{R}^d)),\\ &A(\partial_x)\big((I-\varepsilon A(\partial_x))^{-1}\varphi\big)=(I-\varepsilon A(\partial_x))^{-1}A(\partial_x)\varphi\in L^2([0,T],H^1(\mathbb{R}^d)).\end{aligned}\tag{5.4.13}$$

This implies that (5.4.12) holds also for $\psi \in L^2([0,T] \times \mathbb{R}^d)$, and that $(I - \varepsilon A(\partial_x))^{-1}\varphi$ is an admissible test function in the equation $\partial_t w - A(\partial_x)w = 0$. By these two facts we obtain

$$
\begin{aligned}
0 &= \int_0^T \int_{\mathbb{R}^d} w \left[(\partial_t + A(\partial_x))(I - \varepsilon A(\partial_x))^{-1}\varphi\right] dx\, dt \\
&= \int_0^T \int_{\mathbb{R}^d} w \left[(I - \varepsilon A(\partial_x))^{-1}(\partial_t + A(\partial_x))\varphi\right] dx\, dt \\
&- \varepsilon \int_0^T \int_{\mathbb{R}^d} \sum_{ij} (\partial_t a_{ij})\partial_i \left((I - \varepsilon A(\partial_x))^{-1} w\right) \partial_j \left((I - \varepsilon A(\partial_x))^{-1}\varphi\right) dx\, dt \\
&= \int_0^T \int_{\mathbb{R}^d} w_\varepsilon \left[(\partial_t + A(\partial_x))\varphi\right] dx\, dt - \varepsilon \int_0^T \int_{\mathbb{R}^d} \sum_{ij} (\partial_t a_{ij})\partial_i w_\varepsilon \partial_j \left((I - \varepsilon A(\partial_x))^{-1}\varphi\right) dx\, dt,
\end{aligned}
$$

which exactly means that

$$
\partial_t w_\varepsilon - A(\partial_x)w_\varepsilon = \varepsilon(I - \varepsilon A(\partial_x))^{-1} \sum_{ij} \partial_j(\partial_t a_{ij}\partial_i w^\varepsilon)
$$

in the distributional sense. □

Proof of Proposition 5.4.4. (a) Arguing as in the the first part of the proof of Theorem 5.4.3, with the same notation we have

$$
\begin{aligned}
0 &= \int_0^T \int_{\mathbb{R}^d} \Big(\partial_t v + \sum_i \partial_i(\tilde{b}_i v) - \frac{1}{2}\sum_{ij} \partial_i(a_{ij}\partial_j v) + \lambda v\Big) v^- \, dx\, dt \\
&= \frac{1}{2}\int_0^T \int_{\mathbb{R}^d} \Big[-\frac{d}{dt}(v^-)^2 - \sum_i \tilde{b}_i \partial_i\big((v^-)^2\big) - \sum_{ij} a_{ij}\partial_i v^- \partial_j v^- - 2\lambda (v^-)^2\Big] dx \\
&\le -\frac{1}{2}\int_{\mathbb{R}^d} (v^-)^2(T,x)\, dx - \Big(\lambda - \frac{1}{2}\|(\sum_i \partial_i \tilde{b}_i)^-\|_\infty\Big) \int_0^T \int_{\mathbb{R}^d} (v^-)^2 \, dx\, dt \\
&\le -\Big(\lambda - \frac{1}{2}\|(\sum_i \partial_i \tilde{b}_i)^-\|_\infty\Big) \int_0^T \int_{\mathbb{R}^d} (v^-)^2 \, dx\, dt,
\end{aligned}
$$

and then $v^- = 0$.
(b) It suffices to observe that the above argument works for every $v \in Y$ such that $v(0) \ge 0$ and

$$
\partial_t v + \sum_i \partial_i(\tilde{b}_i v) - \frac{1}{2}\sum_{ij} \partial_i(a_{ij}\partial_j v) \ge 0.
$$

Applying this remark to the function $v := \|u_0\|_{L^\infty(\mathbb{R}^d)} - ue^{-\lambda t}$ with $\lambda > \|(\sum_i \partial_i \tilde{b}_i)^-\|_\infty$, and then letting $\lambda \to \|(\sum_i \partial_i \tilde{b}_i)^-\|_\infty$, the thesis follows.
(c) The argument we use here is reminiscent of the one that we will use in the next paragraph for renormalized solutions. Indeed, in order to prove the thesis, we will implicitly prove that, if $u \in L^2([0,T], H^1(\mathbb{R}^d))$ is a solution of (5.4.2), it is also a renormalized solution (see Definition 5.4.9).
Let us define

$$\beta_\varepsilon(s) := \left(\sqrt{s^2+\varepsilon^2} - \varepsilon\right) \in C^2(\mathbb{R}).$$

Notice that β_ε is convex and

$$\beta_\varepsilon(s) \to |s| \text{ as } \varepsilon \to 0, \quad \beta_\varepsilon(s) - s\beta_\varepsilon'(s) \in [-\varepsilon, 0].$$

Moreover, since $\beta_\varepsilon', \beta_\varepsilon'' \in W^{1,\infty}(\mathbb{R})$, it is easily seen that

$$u \in L^2([0,T], H^1(\mathbb{R}^d)) \quad \Rightarrow \quad \beta_\varepsilon(u), \beta_\varepsilon'(u) \in L^2([0,T], H^1(\mathbb{R}^d)).$$

Fix now a non-negative cut-off function $\varphi \in C_c^\infty(\mathbb{R}^d)$ with $\mathrm{supp}(\varphi) \subset B_2(0)$, and $\varphi = 1$ in $B_1(0)$, and consider the functions $\varphi_R(x) := \varphi(\frac{x}{R})$ for $R \geq 1$.
Thus, since $\beta_\varepsilon'' \geq 0$ and a_{ij} is positive definite, recalling that $\tilde{b}_i = b_i - \frac{1}{2}\sum_j \partial_j a_{ij}$, for any $t \in [0,T]$ we have

$$\begin{aligned}
0 &= \int_0^t \int_{\mathbb{R}^d} \Big(\partial_t u + \sum_i \partial_i(\tilde{b}_i u) - \frac{1}{2}\sum_{ij} \partial_i(a_{ij}\partial_j u)\Big)\beta_\varepsilon'(u)\varphi_R \, dx\, ds \\
&= \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} \bigg(\frac{d}{dt}(\varphi_R \beta_\varepsilon(u)) - 2\sum_i \tilde{b}_i \partial_i(u\beta_\varepsilon'(u)\varphi_R) + 2\sum_i \tilde{b}_i \partial_i(\beta_\varepsilon(u))\varphi_R \\
&\qquad + \sum_{ij} a_{ij}\partial_i u \partial_j u \beta_\varepsilon''(u)\varphi_R + \sum_{ij} a_{ij}\partial_i(\beta_\varepsilon(u))\partial_j\varphi_R\bigg)\, dx\, ds \\
&\geq \frac{1}{2}\int_{\mathbb{R}^d} \varphi_R \beta_\varepsilon(u(t))\, dx - \frac{1}{2}\int_{\mathbb{R}^d} \varphi_R\beta_\varepsilon(u(0))\, dx \\
&\qquad - \int_0^t \int_{\mathbb{R}^d} \sum_i \tilde{b}_i \Big(\partial_i\big((u\beta_\varepsilon'(u) - \beta_\varepsilon(u))\varphi_R\big) + \beta_\varepsilon(u)\partial_i\varphi_R\Big)\, dx\, ds \\
&\qquad - \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} \sum_{ij} \Big((\partial_j a_{ij})\partial_i\varphi_R + a_{ij}\partial_{ij}\varphi_R\Big)\beta_\varepsilon(u)\, dx\, ds \\
&\geq \frac{1}{2}\int_{\mathbb{R}^d} \varphi_R\beta_\varepsilon(u(t))\, dx - \frac{1}{2}\int_{\mathbb{R}^d}\varphi_R\beta_\varepsilon(u(0))\, dx - \int_0^t\int_{\mathbb{R}^d} (\sum_i \partial_i\tilde{b}_i)^-(u\beta_\varepsilon'(u) - \beta_\varepsilon(u))\varphi_R\, dx\, ds \\
&\qquad - \int_0^t \int_{\mathbb{R}^d} \Big(\sum_i b_i\partial_i\varphi_R + \frac{1}{2}\sum_{ij} a_{ij}\partial_{ij}\varphi_R\Big)\beta_\varepsilon(u)\, dx\, ds.
\end{aligned}$$

Observing that $|\beta_\varepsilon(u)| \le |u|$, and using Hölder inequality and the inequalities

$$\frac{1}{R}\chi_{\{R\le|x|\le 2R\}} \le \frac{3}{1+|x|}\chi_{\{|x|\ge R\}}, \qquad \frac{1}{R^2}\chi_{\{R\le|x|\le 2R\}} \le \frac{5}{1+|x|^2}\chi_{\{|x|\ge R\}}, \tag{5.4.14}$$

we get

$$\int_{\mathbb{R}^d}\varphi_R\beta_\varepsilon(u(t))\,dx \le \int_{\mathbb{R}^d}\varphi_R\beta_\varepsilon(u(0))\,dx + 2\varepsilon\int_0^t\int_{|x|\le 2R}(\sum_i \partial_i\tilde{b}_i)^-\,dx\,ds$$
$$+\|\varphi\|_{C^2}\left(6\left\|\frac{b}{1+|x|}\right\|_{L^2([0,T]\times\{|x|\ge R\})}+5\left\|\frac{a}{1+|x|^2}\right\|_{L^2([0,T]\times\{|x|\ge R\})}\right)\|u\|_{L^2([0,t]\times\mathbb{R}^d)}.$$

Letting first $\varepsilon\to 0$ and then $R\to\infty$, we obtain

$$\|u(t)\|_{L^1(\mathbb{R}^d)} \le \|u(0)\|_{L^1(\mathbb{R}^d)} \quad \forall t\in[0,T].$$

□

5.4.3 Existence and uniqueness in the degenerate parabolic case

We now want to drop the uniform ellipticity assumption on a. In this case, to prove existence and uniqueness in $\mathscr{L}_+$, we will need to assume a independent of the space variables.

• **Uniqueness in $\mathscr{L}$**

The uniqueness result is a consequence of the following comparison principle in $\mathscr{L}$ (recall that the comparison principle in said to hold if the inequality between two solutions at time 0 is preserved at later times).

Theorem 5.4.8 (Comparison principle in $\mathscr{L}$). *Let us assume that $a:[0,T]\to\mathcal{S}_+(\mathbb{R}^d)$ and $b:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ are such that:*

1. *$b\in L^1([0,T],BV_{loc}(\mathbb{R}^d,\mathbb{R}^d))$, $\sum_i\partial_i b_i\in L^1_{loc}([0,T]\times\mathbb{R}^d)$;*
2. *$a\in L^\infty([0,T],\mathcal{S}_+(\mathbb{R}^d))$.*

Then (5.4.1) satisfies the comparison principle in $L^1(\mathbb{R}^d)\cap L^\infty(\mathbb{R}^d)$. In particular solutions of the PDE in $\mathscr{L}$, if they exist, are unique.

Since we do not assume any ellipticity of the PDE, in order to prove the above result we use the technique of renormalized solutions, which was first introduced in the study of the Boltzmann equation by DiPerna and P.-L.Lions [54, 55], and then applied in the context of transport equations by many authors (see for example [56, 28, 47, 48, 4]).

Definition 5.4.9. Let $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$, $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ be such that:

1. $b, \sum_i \partial_i b_i \in L^1_{loc}([0,T] \times \mathbb{R}^d)$;
2. $a, \sum_j \partial_j a_{ij}, \sum_{ij} \partial_{ij} a_{ij} \in L^1_{loc}([0,T] \times \mathbb{R}^d)$.

Let $u \in L^\infty_{loc}([0,T] \times \mathbb{R}^d)$ and assume that

$$c := \partial_t u + \sum_i b_i \partial_i u - \frac{1}{2} \sum_{ij} a_{ij} \partial_{ij} u \in L^1_{loc}([0,T] \times \mathbb{R}^d). \tag{5.4.15}$$

We say that u is a renormalized solution of (5.4.15) if, for any convex function $\beta : \mathbb{R} \to \mathbb{R}$ of class C^2, we have

$$\partial_t \beta(u) + \sum_i b_i \partial_i \beta(u) - \frac{1}{2} \sum_{ij} a_{ij} \partial_{ij} \beta(u) \leq c \beta'(u).$$

Equivalently the definition could be given in a partially conservative form:

$$\partial_t \beta(u) + \sum_i \partial_i (b_i \beta(u)) - \frac{1}{2} \sum_{ij} a_{ij} \partial_{ij} \beta(u) \leq c \beta'(u) + (\sum_i \partial_i b_i) \beta(u).$$

Recalling that a is non-negative definite and β is convex, it is simple to check that, if everything is smooth so that one can apply the standard chain rule, every solution of (5.4.15) is a renormalized solution. Indeed, in that case, one gets

$$\partial_t \beta(u) + \sum_i b_i \partial_i \beta(u) - \frac{1}{2} \sum_{ij} a_{ij} \partial_{ij} \beta(u) = c \beta'(u) - \frac{1}{2} \beta''(u) \sum_{ij} a_{ij} \partial_i u \partial_j u \leq c \beta'(u).$$

In our case, a solution of the Fokker-Planck equation is renormalized if

$$\partial_t \beta(u) + \sum_i (b_i - \sum_j \partial_j a_{ij}) \partial_i \beta(u) - \frac{1}{2} \sum_{ij} a_{ij} \partial_{ij} \beta(u) \leq (\frac{1}{2} \sum_{ij} \partial_{ij} a_{ij} - \sum_i \partial_i b_i) u \beta'(u),$$

or equivalently, writing everything in the partially conservative form,

$$\begin{aligned}
\partial_t \beta(u) &+ \sum_i \partial_i ((b_i - \sum_j \partial_j a_{ij}) \beta(u)) - \frac{1}{2} \sum_{ij} a_{ij} \partial_{ij} \beta(u) \\
&\leq (\frac{1}{2} \sum_{ij} \partial_{ij} a_{ij} - \sum_i \partial_i b_i) u \beta'(u) + \sum_i \partial_i (b_i - \sum_j \partial_j a_{ij}) \beta(u) \\
&= (\sum_i \partial_i b_i - \frac{1}{2} \sum_{ij} \partial_{ij} a_{ij})(\beta(u) - u \beta'(u)) - \frac{1}{2} (\sum_{ij} \partial_{ij} a_{ij}) \beta(u).
\end{aligned}$$

Now, since

$$\begin{aligned}\sum_{ij} a_{ij}\partial_{ij}\beta(u) &= \sum_{ij}\partial_j(a_{ij}\partial_i\beta(u)) - \sum_{ij}\partial_j a_{ij}\partial_i\beta(u)\\ &= \sum_{ij}\partial_{ij}(a_{ij}\beta(u)) - 2\sum_{ij}\partial_i((\partial_j a_{ij})\beta(u)) + (\sum_{ij}\partial_{ij}a_{ij})\beta(u),\end{aligned}$$

the above expression can be simplified, and we obtain that a solution of the Fokker-Planck equation is renormalized if and only if

$$\partial_t\beta(u) + \sum_i \partial_i(b_i\beta(u)) - \frac{1}{2}\sum_{ij}\partial_{ij}(a_{ij}\beta(u)) \le (\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij})(\beta(u) - u\beta'(u)). \tag{5.4.16}$$

It is not difficult to prove the following:

Lemma 5.4.10. *Assume that there exist $p, q \in [1,\infty]$ such that*

$$\frac{a}{1+|x|^2} \in L^1([0,T], L^p(\mathbb{R}^d)), \qquad \frac{b}{1+|x|} \in L^1([0,T], L^q(\mathbb{R}^d)),$$

and that

$$(\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij})^- \in L^1_{loc}([0,T]\times\mathbb{R}^d).$$

Setting $a, b = 0$ for $t < 0$, assume moreover that any solution $u \in \mathscr{L}$ of the Fokker-Planck equation in $(-\infty, T)\times\mathbb{R}^d$ is renormalized. Then the comparison principle holds in $\mathscr{L}$.

Proof. By the linearity of the equation, it suffices to prove that

$$u_0 \le 0 \quad \Rightarrow \quad u(t) \le 0 \quad \forall t \in [0,T].$$

Fix a non-negative cut-off function $\varphi \in C_c^\infty(\mathbb{R}^d)$ with $\mathrm{supp}(\varphi) \subset B_2(0)$, and $\varphi = 1$ in $B_1(0)$, and take as renormalization function

$$\beta_\varepsilon(s) := \frac{1}{2}\left(\sqrt{s^2+\varepsilon^2} + s - \varepsilon\right) \in C^2(\mathbb{R}).$$

Notice that β_ε is convex and

$$\beta_\varepsilon(s) \to s^+ \text{ as } \varepsilon \to 0, \quad \beta_\varepsilon(s) - s\beta_\varepsilon'(s) \in [-\varepsilon, 0].$$

By (5.4.16), we know that

$$\partial_t\beta_\varepsilon(u) + \sum_i \partial_i(b_i\beta_\varepsilon(u)) - \frac{1}{2}\sum_{ij}\partial_{ij}(a_{ij}\beta_\varepsilon(u)) \le (\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij})(\beta_\varepsilon(u) - u\beta_\varepsilon'(u))$$

in the sense of distributions in $(-\infty, T) \times \mathbb{R}^d$. Using as test function $\varphi_R(x) := \varphi(\frac{x}{R})$ for $R \geq 1$, we get

$$\frac{d}{dt}\int_{\mathbb{R}^d} \varphi_R \beta_\varepsilon(u)\,dx \leq \int_{\mathbb{R}^d}\Big(\sum_i b_i(t)\partial_i\varphi_R + \frac{1}{2}\sum_{ij} a_{ij}(t)\partial_{ij}\varphi_R\Big)\beta_\varepsilon(u)\,dx$$
$$+ \int_{\mathbb{R}^d}\varphi_R\Big(\sum_i \partial_i b_i(t) - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij}(t)\Big)(\beta_\varepsilon(u) - u\beta'_\varepsilon(u))\,dx$$

Observing that $|\beta_\varepsilon(u)| \leq |u|$, by Hölder inequality and the inequalities (5.4.14) we can bound the first integral in the right hand side, uniformly with respect to ε, with

$$\|\varphi\|_{C^2}\int_{\{|x|\geq R\}}\left(3\frac{|b(t,x)|}{1+|x|} + \frac{5}{2}\frac{|a(t,x)|}{(1+|x|^2)}\right)|u(t,x)|\,dx$$
$$\leq \|\varphi\|_{C^2}\left(3\Big\|\frac{b(t)}{1+|x|}\Big\|_{L^p(\{|x|\geq R\})}\|u(t)\|_{L^{p'}(\mathbb{R}^d)}\right.$$
$$\left. + \frac{5}{2}\Big\|\frac{a(t)}{1+|x|^2}\Big\|_{L^q(\{|x|\geq R\})}\|u(t)\|_{L^{q'}(\mathbb{R}^d)}\right)$$

(recall that $u \in \mathscr{L}$, and thus $u \in L^\infty([0,T], L^r(\mathbb{R}^d))$ for any $r \in [1,\infty]$), while the second integral is bounded by

$$\varepsilon\int_{\{|x|\leq 2R\}}\Big(\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij}\Big)^-\,dx.$$

Letting first $\varepsilon \to 0$ and then $R \to \infty$, we get

$$\frac{d}{dt}\int_{\mathbb{R}^d} u^+\,dx \leq 0$$

in the sense of distribution in $(-\infty, T)$. Since the function vanishes for negative times, we conclude $u^+ = 0$. □

Now Theorem 5.4.8 is a direct consequence of the following:

Proposition 5.4.11. *Let us assume that $a : [0,T] \to \mathcal{S}_+(\mathbb{R}^d)$ and $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ are such that:*

1. *$b \in L^1([0,T], BV_{loc}(\mathbb{R}^d, \mathbb{R}^d))$, $\sum_i \partial_i b_i \in L^1_{loc}([0,T] \times \mathbb{R}^d)$;*
2. *$a \in L^\infty([0,T], \mathcal{S}_+(\mathbb{R}^d))$*

Then any distributional solution $u \in L^\infty_{loc}([0,T] \times \mathbb{R}^d)$ of (5.4.15) is renormalized.

Proof. We take η, a smooth convolution kernel in $\mathbb{R}^d$, and we mollify the equation with respect to the spatial variable obtaining

$$\partial_t u^\varepsilon + \sum_i b_i \partial_i u^\varepsilon - \frac{1}{2}\sum_{ij} a_{ij}\partial_{ij} u^\varepsilon = c * \eta_\varepsilon - r^\varepsilon, \tag{5.4.17}$$

where

$$r^\varepsilon := \sum_i (b_i \partial_i u) * \eta_\varepsilon - \sum_i b_i \partial_i (u * \eta_\varepsilon), \quad u^\varepsilon := u * \eta_\varepsilon.$$

By the smoothness of u^ε with respect to x, by (5.4.17) we have that $\partial_t u^\varepsilon \in L^1_{loc}$. Thus by the standard chain rule in Sobolev spaces we get that u^ε is a renormalized solution, that is

$$\partial_t \beta(u^\varepsilon) + \sum_i b_i \partial_i \beta(u^\varepsilon) - \frac{1}{2}\sum_{ij} a_{ij}\partial_{ij}\beta(u^\varepsilon) \le (c * \eta_\varepsilon - r^\varepsilon)\beta'(u^\varepsilon)$$

for any $\beta \in C^2(\mathbb{R})$ convex. Passing to the limit in the distributional sense as $\varepsilon \to 0$ in the above identity, the convergence of all the terms is trivial except for $r^\varepsilon \beta'(u^\varepsilon)$.
Let σ_η be any weak limit point of $r^\varepsilon\beta'(u^\varepsilon)$ in the sense of measures (such a cluster point exists since $r^\varepsilon\beta'(u^\varepsilon)$ is bounded in L^1_{loc}). Thus we get

$$\partial_t \beta(u) + \sum_i b_i \partial_i \beta(u) - \frac{1}{2}\sum_{ij} a_{ij}\partial_{ij}\beta(u) - c\beta'(u) \le -\sigma_\eta \le |\sigma_\eta|.$$

Since the left hand side is independent of η, in order to conclude the proof it suffices to prove that $\bigwedge_\eta |\sigma_\eta| = 0$, where η varies in a dense countable set of convolution kernels. This fact is implicitly proved in [5, Theorem 34], see in particular Step 3 therein. □

• **Existence in $\mathscr{L}_+$**
We can now prove an existence and uniqueness result in the class $\mathscr{L}_+$.

Theorem 5.4.12. *Let us assume that $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}(\mathbb{R}^d)$ and $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ are bounded functions such that*

$$(\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij} a_{ij})^- \in L^1([0,T], L^\infty(\mathbb{R}^d)).$$

Then, for any $\mu_0 = \rho_0 \mathscr{L}^d \in \mathcal{M}_+(\mathbb{R}^d)$, with $\rho_0 \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$, there exists a solution of (5.1.2) in $\mathscr{L}_+$. If moreover $b \in L^1([0,T], BV_{loc}(\mathbb{R}^d))$, $\sum_i \partial_i b_i \in L^1_{loc}([0,T] \times \mathbb{R}^d)$, and a is independent of x, then this solution turns out to be unique.

Proof. **Existence:** it suffices to approximate the coefficients a and b locally uniformly with smooth uniformly bounded coefficients a^n and b^n such that $(\sum_i \partial_i b_i^n - \frac{1}{2}\sum_{ij}\partial_{ij} a_{ij}^n)^-$

is uniformly bounded in $L^1([0,T],L^\infty(\mathbb{R}^d))$. Indeed, if we now consider the approximate solutions $\mu_t^n = \rho_t^n \mathscr{L}^d \in \mathcal{M}_+(\mathbb{R}^d)$, we know that

$$\partial_t \rho_t^n + \sum_i \partial_i (b_i^n \rho_t^n) - \frac{1}{2}\sum_{ij} \partial_{ij}(a_{ij}^n \rho_t^n) = 0,$$

that is

$$\partial_t \rho_t^n - \frac{1}{2} a_{ij}^n \partial_{ij} \rho_t^n + \sum_i (b_i^n - \sum_j \partial_j a_{ij}^n)\partial_i \rho_t^n + (\sum_i \partial_i b_i^n - \frac{1}{2}\sum_{ij} \partial_{ij} a_{ij}^n)\rho_t^n = 0.$$

Using the Feynman-Kac's formula, we obtain the bound

$$\|\rho_t^n\|_{L^\infty(\mathbb{R}^d)} \le \|\rho_0\|_{L^\infty(\mathbb{R}^d)} e^{\int_0^t \|(\sum_i \partial_i b_i^n(s) - \frac{1}{2}\sum_{ij}\partial_{ij} a_{ij}^n(s))^-\|_{L^\infty(\mathbb{R}^d)}\,dt}.$$

So we see that the approximate solutions are non-negative and uniformly bounded in $L^1 \cap L^\infty$ (the bound in L^1 follows by the constancy of the map $t \mapsto \|\rho_t^n\|_{L^1}$ (observe that $\rho_t^n \ge 0$ and recall Remark 5.2.8)). Therefore, any weak limit is a solution of the PDE in $\mathscr{L}_+$.
Uniqueness: it follows by Theorem 5.4.8. □

5.5 Conclusions

Let us now combine the results proved in Sections 5.2 and 5.4 in order to get existence and uniqueness of SLF. The first theorem follows directly by Corollary 5.3.6 and Theorem 5.1.3, while the second is a consequence of Corollary 5.3.6 and Theorem 5.1.4.

Theorem 5.5.1. *Let us assume that $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ and $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ are bounded functions such that:*

1. $\sum_j \partial_j a_{ij} \in L^\infty([0,T] \times \mathbb{R}^d)$ *for* $i = 1,\dots,d$,
2. $\partial_t a_{ij} \in L^\infty([0,T] \times \mathbb{R}^d)$ *for* $i,j = 1,\dots,d$;
3. $(\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij} a_{ij})^- \in L^\infty([0,T] \times \mathbb{R}^d)$;
4. $\langle \xi, a(t,x)\xi\rangle \ge \alpha|\xi|^2 \ \forall (t,x) \in [0,T] \times \mathbb{R}^d$, *for some* $\alpha > 0$;
5. $\frac{a}{1+|x|^2} \in L^2([0,T] \times \mathbb{R}^d)$, $\frac{b}{1+|x|} \in L^2([0,T] \times \mathbb{R}^d)$.

Then there exists a unique SLF (in the sense of Corollary 5.3.6).

If moreover $(b^n, a^n) \to (b, a)$ *in* $L^1_{loc}([0,T] \times \mathbb{R}^d)$ *and* $(\sum_i \partial_i b_i^n - \frac{1}{2}\sum_{ij} \partial_{ij} a_{ij}^n)^-$ *are uniformly bounded in* $L^1([0,T], L^\infty(\mathbb{R}^d))$*, then the Feynman- Kac formula implies (ii) of Theorem 5.3.7 (see the proof of Theorem 5.4.12). Thus we have stability of SLF.*

Theorem 5.5.2. *Let us assume that* $a : [0,T] \to \mathcal{S}(\mathbb{R}^d)$ *and* $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ *are bounded functions such that:*

1. $b \in L^1([0,T], BV_{loc}(\mathbb{R}^d))$, $\sum_i \partial_i b_i \in L^1_{loc}([0,T] \times \mathbb{R}^d)$;
2. $(\sum_i \partial_i b_i)^- \in L^1([0,T], L^\infty(\mathbb{R}^d))$.

Then there exists a unique SLF (in the sense of Corollary 5.3.6).

If moreover $(b^n, a^n) \to (b, a)$ *in* $L^1_{loc}([0,T] \times \mathbb{R}^d)$ *and* $(\sum_i \partial_i b_i^n - \frac{1}{2}\sum_{ij} \partial_{ij} a_{ij}^n)^-$ *are uniformly bounded in* $L^1([0,T], L^\infty(\mathbb{R}^d))$*, then the Feynman-Kac formula implies (ii) of Theorem 5.3.7 (see the proof of Theorem 5.4.12). Thus we have stability of SLF.*

In particular, by Corollary 5.3.9 and the Feynman-Kac formula (see the proof of Theorem 5.4.12), the following vanishing viscosity result for RLF holds:

Theorem 5.5.3. *Let us assume that* $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ *is bounded and:*

1. $b \in L^1([0,T], BV_{loc}(\mathbb{R}^d))$, $\sum_i \partial_i b_i \in L^1_{loc}([0,T] \times \mathbb{R}^d)$;
2. $(\sum_i \partial_i b_i)^- \in L^1([0,T], L^\infty(\mathbb{R}^d))$.

Let $\{\nu_x^\varepsilon\}_{x \in \mathbb{R}^d}$ *be the unique SLF relative to* $(b, \varepsilon I)$*, with* $\varepsilon > 0$*, and* $\{\nu_x\}_{x \in \mathbb{R}^d}$ *be the RLF relative to* $(b, 0)$ *(which is uniquely determined* $\mathscr{L}^d$*-a.e. by the results in [4]). Then, as* $\varepsilon \to 0$*,*

$$\int_{\mathbb{R}^d} \nu_x^\varepsilon f(x)\, dx \rightharpoonup^* \int_{\mathbb{R}^d} \nu_x f(x)\, dx \quad \text{in } \mathcal{M}(\Gamma_T) \text{ for any } f \in C_c(\mathbb{R}^d).$$

We finally combine an important uniqueness result of Stroock and Varadhan (see Theorem 5.2.2) with the well-posedness results on Fokker-Planck of the previous section. By Theorem 5.2.2, Lemma 5.2.3 applied with $A = \mathbb{R}^d$ and Corollary 5.4.5, we have:

Theorem 5.5.4. *Let us assume that* $a : [0,T] \times \mathbb{R}^d \to \mathcal{S}_+(\mathbb{R}^d)$ *and* $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ *are bounded functions such that:*

1. $\langle \xi, a(t,x)\xi \rangle \geq \alpha |\xi|^2 \ \forall (t,x) \in [0,T] \times \mathbb{R}^d$*, for some* $\alpha > 0$*;*
2. $|b(t,x) - b(s,y)| + \|a(t,x) - a(s,y)\| \leq C\,(|x-y|^\gamma + |t-s|^\gamma)\ \forall (t,x), (s,y) \in [0,T] \times \mathbb{R}^d$*, for some* $\gamma \in (0,1]$*,* $C \geq 0$*;*

3. *$\sum_j \partial_j a_{ij} \in L^\infty([0,T]\times\mathbb{R}^d)$ for $i=1,\dots,d$, $(\sum_i \partial_i b_i - \frac{1}{2}\sum_{ij}\partial_{ij}a_{ij})^- \in L^\infty([0,T]\times\mathbb{R}^d)$;*

4. *$\frac{a}{1+|x|^2} \in L^2([0,T]\times\mathbb{R}^d)$, $\frac{b}{1+|x|} \in L^2([0,T]\times\mathbb{R}^d)$.*

Then, there exists a unique martingale solution starting from x (at time 0) for any $x \in \mathbb{R}^d$.

We remark that this result is not interesting by itself, since it can be proved that the martingale problem starting from any $x \in \mathbb{R}^d$ at any initial time $s \in [0,T]$ is well-posed also under weaker regularity assumptions (see [125, Chapters 6 and 7]). We stated it just because we believe that it is an interesting example of how existence and uniqueness at the PDE level can be combined with a refined analysis at the level of the uniqueness of martingale solutions. It is indeed in this spirit that we generalize Theorem 5.2.2 in the following section, hoping that it could be useful for further analogous applications.

5.6 A generalized uniqueness result for martingale solutions

Here we generalize Theorem 5.2.2, using the notation introduced in Paragraph 5.3.1.

Proposition 5.6.1. *For any $(s,x) \in [0,T]\times\mathbb{R}^d$, let $C_{x,s}$ be a subset of martingale solutions of the SDE starting from x at time s, and let us make the following assumptions: there exists a measure $\mu_0 \in \mathcal{M}_+(\mathbb{R}^d)$ such that:*

(i) $\forall s \in [0,T]$, $C_{x,s}$ is convex for μ_0-a.e. x;

(ii) $\forall s \in [0,T]$, $\forall t \in [s,T]$,

$$\text{for } \mu_0\text{-a.e. } x, \qquad (e_t)_\# \nu^1_{x,s} = (e_t)_\# \nu^2_{x,s} \quad \forall \nu^1_{x,s}, \nu^2_{x,s} \in C_{x,s};$$

(iii) for μ_0-a.e. x, for any $\nu_x \in C_x := C_{x,0}$, for ν_x-a.e. γ,

$$\forall t \in [0,T], \qquad \nu^{i,\gamma}_{x,\mathcal{F}_t} := (\nu^i_x)^\gamma_{\mathcal{F}_t} \in C_{\gamma(t),t},$$

where, with the above notation, we mean that the restriction of $\nu^{i,\gamma}_{x,\mathcal{F}_t}$ to $\Gamma^t_T := C([t,T],\mathbb{R}^d)$ is a martingale solution starting from $\gamma(t)$ at time t;

(iv) the solution of (5.1.2) starting from μ_0 given by $\mu_t := (e_t)_\# \int_{\mathbb{R}^d} \nu^1_x \, d\mu_0(x)$ for a measurable selections $\{\nu_x\}_{x\in\mathbb{R}^d}$ with $\nu_x \in C_x$ (observe that μ_t does not depends on the choice of $\nu_x \in C_x$ by (ii)), satisfies $\mu_t \ll \mu_0$ for any $t \in [0,T]$.

Then, given two measurable families of probability measures $\{\nu_x^1\}_{x\in\mathbb{R}^d}$ and $\{\nu_x^2\}_{x\in\mathbb{R}^d}$ with $\nu_x^1,\nu_x^2 \in C_x$, $\nu_x^1 = \nu_x^2$ for μ_0-a.e. x. In particular, by standard measurable selection theorems (see for instance [125, Chapter 12]), C_x is a singleton for μ_0-a.e. x.

Proof. Let $\{\nu_x^1\}_{x\in\mathbb{R}^d}$ and $\{\nu_x^2\}_{x\in\mathbb{R}^d}$ be two measurable families of probability measures with $\nu_x^1,\nu_x^2 \in C_x$, and fix $0 < t_1 < \ldots < t_n \le T$.
Claim: for μ_0-a.e. x, for ν_x^i-a.e. γ $(i = 1,2)$,

$$\nu^{i,\tilde\gamma}_{x,\mathcal{F}_{t_n}} \in C_{\tilde\gamma(t_n),t_n} \quad \text{for } \nu^{i,\gamma}_{x,M^{t_1,\ldots,t_n}}\text{-a.e. } \tilde\gamma$$

where $\nu^{i,\gamma}_{x,M^{t_1,\ldots,t_n}} := (\nu_x^i)^\gamma_{M^{t_1,\ldots,t_n}}$.
This claim follows observing that, by assumption (iii), for μ_0-a.e. x there exists a subset $\Gamma_x \subset \Gamma_T$ such that $\nu_x^i(\Gamma_x) = 1$ and $\nu^{i,\gamma}_{x,\mathcal{F}_{t_n}} \in C_{\gamma(t_n),t_n}$ for any $\gamma \in \Gamma_x$. Thus, by (5.3.1) applied with $\nu := \nu_x^i$, $A := \Gamma_T$, $B := \Gamma_x$, and with $M^{t_1,\ldots,t_n}$ in place of $\mathcal{F}_{t_n}$, one obtains

$$0 = \nu_x^i(\Gamma_x^c) = \int_{\Gamma_T} \nu^{i,\gamma}_{x,M^{t_1,\ldots,t_n}}(\Gamma_x^c)\, d\nu_x^i(\gamma),$$

that is,

$$\text{for } \nu_x^i\text{-a.e. } \gamma, \qquad \nu^{i,\gamma}_{x,M^{t_1,\ldots,t_n}}(\Gamma_x) = 1.$$

This, together with assumption (iii), implies the claim.

By (5.3.3), $\nu^{i,\gamma}_{x,M^{t_1,\ldots,t_n}}$ is concentrated on the set $\{\tilde\gamma \mid \tilde\gamma(t_n) = \gamma(t_n)\}$, and so, by the claim above, we get

$$\nu^{i,\tilde\gamma}_{x,\mathcal{F}_{t_n}} \in C_{\gamma(t_n),t_n} \quad \text{for } \nu^{i,\gamma}_{x,M^{t_1,\ldots,t_n}}\text{-a.e. } \tilde\gamma.$$

Let $A \subset \mathbb{R}^d$ be such that $\mu_0(A^c) = 0$ and assumption (i) is true for any $x \in A$. By assumption (iv), we have $\mu_{t_n}(A^c) = 0 = \int_{\mathbb{R}^d\times\Gamma_T} 1_{A^c}(\gamma(t_n))\, d\nu_x^i(\gamma)\, d\mu_0(x)$, that is

$$\text{for } \mu_0\text{-a.e. } x, \qquad \gamma(t_n) \in A \quad \text{for } \nu_x^i\text{-a.e } \gamma. \tag{5.6.1}$$

Thus, for μ_0-a.e. x, $C_{\gamma(t_n),t_n}$ is convex for ν_x^i-a.e γ, and so, by (5.3.4) applied with ν_x^i, we obtain that

$$\text{for } \mu_0\text{-a.e. } x, \qquad \nu^{i,\gamma}_{x,M^{t_1,\ldots,t_n}} \in C_{\gamma(t_n),t_n} \quad \text{for } \nu_x^i\text{-a.e. } \gamma \tag{5.6.2}$$

(where, with the above notation, we again mean that the restriction of $\nu^{i,\gamma}_{x,M^{t_1,\ldots,t_n}}$ to $\Gamma_T^{t_n}$ is a martingale solution starting from $\gamma(t_n)$ at time t_n). We now want to prove that, for all $n \ge 1$, $0 < t_1 < \ldots < t_n \le T$, we have that, for μ_0-a.e. x,

$$\int_{\Gamma_T} f_1(e_{t_1}(\gamma))\ldots f_n(e_{t_n}(\gamma))\, d\nu_x^1(\gamma) = \int_{\Gamma_T} f_1(e_{t_1}(\gamma))\ldots f_n(e_{t_n}(\gamma))\, d\nu_x^2(\gamma) \tag{5.6.3}$$

for any $f_i \in C_c(\mathbb{R}^d)$. We observe that (5.6.3) is true for $n = 1$ by assumption (ii). We want to prove it for any n by induction. Let us assume (5.6.3) true for $n-1$, and let us

prove it for n.
We want to show that

$$\int_{\Gamma_T} f_1(e_{t_1}(\gamma))\dots f_n(e_{t_n}(\gamma))\,d\nu_x^1(\gamma) = \int_{\Gamma_T} f_1(e_{t_1}(\gamma))\dots f_n(e_{t_n}(\gamma))\,d\nu_x^2(\gamma),$$

which can be written also as

$$\mathbb{E}^{\nu_x^1}\left[f_1(e_{t_1})\dots f_n(e_{t_n})\right] = \mathbb{E}^{\nu_x^2}\left[f_1(e_{t_1})\dots f_n(e_{t_n})\right],$$

where $\mathbb{E}^\nu := \int_{\Gamma_T} d\nu$. Now we observe that, for $i=1,2$,

$$\begin{aligned}\mathbb{E}^{\nu_x^i}\left[f_1(e_{t_1})\dots f_n(e_{t_n})\right] &= \mathbb{E}^{\nu_x^i}\left[\mathbb{E}^{\nu_x^i}[f_1(e_{t_1})\dots f_n(e_{t_n}) \mid M^{t_1,\dots,t_{n-1}}]\right]\\ &= \mathbb{E}^{\nu_x^i}\left[f_1(e_{t_1})\dots f_{n-1}(e_{t_{n-1}})\mathbb{E}^{\nu_x^i}[f_n(e_{t_n}) \mid M^{t_1,\dots,t_{n-1}}]\right]\\ &= \mathbb{E}^{\nu_x^i}\left[f_1(e_{t_1})\dots f_{n-1}(e_{t_{n-1}})\psi_x^i(e_{t_1},\dots,e_{t_{n-1}})\right],\end{aligned}$$

where $\psi_x^i(e_{t_1},\dots,e_{t_{n-1}}) := \mathbb{E}^{\nu_x^i}[f_n(e_{t_n}) \mid M^{t_1,\dots,t_{n-1}}]$. Let $\phi \in C_c(\mathbb{R}^d)$, and let us prove that

$$\begin{aligned}\int_{\mathbb{R}^d} &\mathbb{E}^{\nu_x^1}\left[f_1(e_{t_1})\dots f_{n-1}(e_{t_{n-1}})\psi_x^1(e_{t_1},\dots,e_{t_{n-1}})\right]\phi(x)\,d\mu_0(x)\\ &= \int_{\mathbb{R}^d} \mathbb{E}^{\nu_x^2}\left[f_1(e_{t_1})\dots f_{n-1}(e_{t_{n-1}})\psi_x^2(e_{t_1},\dots,e_{t_{n-1}})\right]\phi(x)\,d\mu_0(x). \quad (5.6.4)\end{aligned}$$

Let $B \subset \mathbb{R}^d$ be such that $\mu_0(B^c)=0$ and assumption (ii') is true for any $x \in B$. By assumption (iv), we also have $\mu_{t_{n-1}}(B^c) = 0 = \int_{\mathbb{R}^d\times\Gamma_T} 1_{B^c}(e_{t_{n-1}}(\gamma))\,d\nu_x^i(\gamma)\,d\mu_0(x)$, that is

$$\text{for } \mu_0\text{-a.e. } x, \qquad \gamma(t_{n-1}) \in B \quad \text{for } \nu_x^i\text{-a.e. } \gamma. \qquad (5.6.5)$$

Let us consider $\nu^{i,\gamma}_{x,M^{t_1,\dots,t_{n-1}}}$. By (5.6.2),

$$\text{for } \mu_0\text{-a.e. } x, \qquad \nu^{i,\gamma}_{x,M^{t_1,\dots,t_{n-1}}} \in C_{\gamma(t_{n-1}),t_{n-1}} \quad \text{for } \nu_x^i\text{-a.e. } \gamma,$$

and, combining this with (5.6.5), we obtain

$$\text{for } \mu_0\text{-a.e. } x, \qquad \nu^{i,\gamma}_{x,M^{t_1,\dots,t_{n-1}}} \in C_{\gamma(t_{n-1}),t_{n-1}} \quad \text{and} \quad \gamma(t_{n-1}) \in B \quad \text{for } \nu_x^i\text{-a.e. } \gamma.$$

By assumption (ii) applied with $t=t_n$, this implies that

$$\text{for } \mu_0\text{-a.e. } x, \qquad (e_{t_n})_\#\nu^{1,\gamma}_{x,M^{t_1,\dots,t_{n-1}}} = (e_{t_n})_\#\nu^{2,\gamma}_{x,M^{t_1,\dots,t_{n-1}}} \quad \text{for } \nu_x^i\text{-a.e. } \gamma,$$

which give us that

$$\text{for } \mu_0\text{-a.e. } x, \qquad \psi^1_x(e_{t_1},\dots,e_{t_{n-1}}) = \psi^2_x(e_{t_1},\dots,e_{t_{n-1}}) \quad \text{for } \nu^i_x\text{-a.e. } \gamma. \tag{5.6.6}$$

Thus we get

$$\begin{aligned}
\int_{\mathbb{R}^d} & \mathbb{E}^{\nu^1_x}\left[f_1(e_{t_1})\dots f_{n-1}(e_{t_{n-1}})\psi^1_x(e_{t_1},\dots,e_{t_{n-1}})\right]\phi(x)\,d\mu_0(x)\\
&= \int_{\mathbb{R}^d} \mathbb{E}^{\nu^2_x}\left[f_1(e_{t_1})\dots f_{n-1}(e_{t_{n-1}})\psi^1_x(e_{t_1},\dots,e_{t_{n-1}})\right]\phi(x)\,d\mu_0(x)\\
&\overset{(5.6.6)}{=} \int_{\mathbb{R}^d} \mathbb{E}^{\nu^2_x}\left[f_1(e_{t_1})\dots f_{n-1}(e_{t_{n-1}})\psi^2_x(e_{t_1},\dots,e_{t_{n-1}})\right]\phi(x)\,d\mu_0(x),
\end{aligned}$$

where the first equality in the above equation follows by the inductive hypothesis. Now, by (5.6.4) and the arbitrariness of ϕ and of f_j, with $j = 1,\dots,n$, we obtain that, for all $n \geq 1$, $0 < t_1 < \dots < t_n \leq T$, we have

$$\text{for } \mu_0\text{-a.e. } x, \qquad (e_{t_1},\dots,e_{t_n})_\#\nu_x = (e_{t_1},\dots,e_{t_n})_\#\tilde{\nu}_x \quad \forall t_1,\dots,t_n \in [0,T].$$

Considering only rational times, we get that there exists a subset $D \subset \mathbb{R}^d$, with $\mu_0(D^c) = 0$, such that, for any $x \in D$,

$$(e_{t_1},\dots,e_{t_n})_\#\nu_x = (e_{t_1},\dots,e_{t_n})_\#\tilde{\nu}_x \quad \text{for any } t_1,\dots,t_n \in [0,T]\cap\mathbb{Q}.$$

By continuity, this implies that, for any $x \in D$, $\nu_x = \tilde{\nu}_x$, as wanted. □

The above result apply, for example, in the case when $C_{x,s}$ denotes the set of all martingale solutions starting from x. In particular, we remark that, by the above proof, one obtains the well-known fact that, if ν_x is a martingale solution starting from x (at time 0), then, for any $0 \leq t_1 \leq \dots \leq t_n \leq T$, $\nu^\gamma_{x,M^{t_1,\dots,t_n}}$ is a martingale solution starting from $\gamma(t_n)$ at time t_n. More in general, since martingale solutions are closed by convex combination, is μ is a probability measure on $\mathbb{R}^d$, the average $\int_{\mathbb{R}^d} \nu^\gamma_{x,M^{t_1,\dots,t_n}}\,d\mu(x)$ is a martingale solution starting from $\gamma(t_n)$ at time t_n.
Observe that assumption (iv) in the above theorem was necessary only to deduce, from a μ_0-a.e. assumption, a μ_t-a.e. property. Thus, the above proof give us the following result:

Proposition 5.6.2. *For any $(s,x) \in [0,T]\times\mathbb{R}^d$, let $C_{x,s}$ be a convex subset of martingale solutions of the SDE starting from x at time s, and let us make the following assumption: there exists a measure $\mu_0 \in \mathcal{M}_+(\mathbb{R}^d)$ such that:*

(i) $\forall t \in [0,T]$, *for μ_0-a.e. x,*

$$(e_t)_\#\nu^1_x = (e_t)_\#\nu^2_x \quad \forall \nu^1_x,\nu^2_x \in C_x := C_{x,0}.$$

If (i) holds, we can define $\mu_t := (e_t)_\# \int_{\mathbb{R}^d} \nu_x \, d\mu_0(x)$ *for a measurable selections* $\{\nu_x\}_{x\in\mathbb{R}^d}$ *with* $\nu_x \in C_x$*, and this definition does not depends on the choice of* $\nu_x \in C_x$*. We now assume that:*

*(i') * $\forall s \in [0,T]$, $\forall t \in [s,T]$, *for* μ_s*-a.e.* x,

$$(e_t)_\# \nu^1_{x,s} = (e_t)_\# \nu^2_{x,s} \quad \forall \nu^1_{x,s}, \nu^2_{x,s} \in C_{x,s};$$

(ii) $\forall s \in [0,T]$, $C_{x,s}$ *is convex for* μ_s*-a.e.* x*;*

(iii) for μ_0*-a.e.* x*, for any* $\nu_x \in C_x$*, for* ν_x*-a.e.* γ*,*

$$\forall t \in [0,T], \qquad \nu^{i,\gamma}_{x,\mathcal{F}_t} := (\nu^i_x)^\gamma_{\mathcal{F}_t} \in C_{\gamma(t),t},$$

where, with the above notation, we mean that the restriction of $\nu^{i,\gamma}_{x,\mathcal{F}_t}$ *to* Γ^t_T *is a martingale solution starting from* $\gamma(t)$ *at time* t*.*

Then, given two measurable families of probability measures $\{\nu^1_x\}_{x\in\mathbb{R}^d}$ *and* $\{\nu^2_x\}_{x\in\mathbb{R}^d}$ *with* $\nu^1_x, \nu^2_x \in C_x$*,* $\nu^1_x = \nu^2_x$ *for* μ_0*-a.e.* x*. In particular, by standard measurable selection theorems (see for instance [125, Chapter 12]),* C_x *is a singleton for* μ_0*-a.e.* x*.*

Chapter 6

Appendix

6.1 Semi-concave functions

We give the definition of semi-concave function and we recall their main properties. The main reference on semi-concave functions is the book [41].

We first recall the definition of a modulus (of continuity).

Definition 6.1.1 (Modulus). A *modulus* ω is a continuous non-decreasing function $\omega : [0, +\infty) \to [0, +\infty)$ such that $\omega(0) = 0$.

We will say that a modulus is *linear* if it is of the form $\omega(t) = kt$, where $k \geq 0$ is some fixed constant.

We will need the notion of superdifferential. We define it in an intrinsic way on a manifold.

Definition 6.1.2 (Superdifferential). Let $f : M \to \mathbb{R}$ be a function. We say that $p \in T_x^*M$ is a *superdifferential* of f at $x \in M$, and we write $p \in D^+f(x)$, if there exists a function $g : V \to \mathbb{R}$, defined on some open subset $U \subset M$ containing x, such that $g \geq f$, $g(x) = f(x)$, and g is differentiable at x with $d_xg = p$.

We now give the definition of a semi-concave function on an open subset of a Euclidean space.

Definition 6.1.3 (Semi-concavity). Let $U \subset \mathbb{R}^n$ open. A function $f : U \to \mathbb{R}$ is said to be *semi-concave* in U with modulus ω (equivalently ω-*semi-concave*) if, for each $x \in U$, we have

$$f(y) - f(x) \leq \langle l_x, y - x \rangle + \|y - x\|\omega(\|y - x\|)$$

for a certain linear form $l_x : \mathbb{R}^n \to \mathbb{R}$.
Note that necessarily $l_x \in D^+f(x)$. Moreover we say that $f : U \to \mathbb{R}$ is *locally semi-concave* if, for each $x \in U$, there exists an open neighborhood of x in which f is semi-concave for a certain modulus.

We will say that the function $f : U \to \mathbb{R}$ is locally semi-concave with a linear modulus if, for each $x \in U$, we can find an open neighborhood V_x such that the restriction $f|_{V_x}$ is ω-semi-concave, with ω a linear modulus.

Proposition 6.1.4. *1) Suppose $f_i : U \to \mathbb{R}, i = 1, \dots, k$ is ω_i-semi-concave, where U is an open subset of $\mathbb{R}^n$. Then we have:*

(i) *for any $\alpha_1, \dots, \alpha_k \geq 0$, the functions $\sum_{i=1}^k \alpha_i f_i$ is $(\sum_{i=1}^k \alpha_i \omega_i)$-semi-concave on U.*

(ii) *the function $\min_{i=1}^k f_i$ is $(\max_{i=1}^k \omega_i)$-semi-concave.*

2) Any C^1 function is locally semi-concave.

Proof. The proof of 1)(i) is obvious. For the proof of (ii), we fix $x \in U$, and we find $i_0 \in \{1, \dots, k\}$ such that $\min_{i=1}^k f_i(x) = f_{i_0}(x)$. Since f_{i_0} is ω_{i_0}-semi-concave, we can find a linear map $l_x : \mathbb{R}^n \to \mathbb{R}$ such that

$$\forall y \in U, \quad f_{i_0}(y) - f_{i_0}(x) \leq l_x(y - x) + \|y - x\| \omega_{i_0}(\|y - x\|).$$

It clearly follows that

$$\forall y \in U, \quad \min_{i=1}^k f_i(y) - \min_{i=1}^k f_i(x) \leq l_x(y - x) + \|y - x\| \max_{i=1}^k \omega_i(\|y - x\|).$$

To prove 2), consider an open convex subset C with $\bar{C}$ compact and contained in U. By compactness of $\bar{C}$ and continuity of $x \mapsto d_x f$, we can find a modulus ω, which is a modulus of continuity for the map $x \mapsto d_x f$ on C. The Mean Value Formula in integral form

$$f(y) - f(x) = \int_0^1 d_{tx+(1-t)y} f(y - x)\, dt,$$

which is valid for every $y, x \in C$ implies that

$$\forall x, y \in U, \quad f(y) - f(x) \leq d_x f(y - x) + \|y - x\| \omega(\|y - x\|).$$

Therefore f is ω-semi-concave in the open subset C. □

We now state and prove the first important consequences of the definition of semi-concavity.

Lemma 6.1.5. *Suppose U is an open subset of $\mathbb{R}^n$. Let $f : U \to \mathbb{R}$ be an ω-semi-concave function. Then we have:*

(i) for every compact subset $K \subset U$, we can find a constant A such that for every $x \in K$, and every linear form l_x on $\mathbb{R}^n$ satisfying

$$\forall y \in U, \quad f(y) - f(x) \leq \langle l_x, y - x\rangle + \|y - x\|\omega(\|y - x\|),$$

we have $\|l_x\| \leq A$;

(ii) the function f is locally Lipschitz.

Proof. From the definition, it follows that a semi-concave function is locally bounded from above. We now show that f is also locally bounded from below. Fix a (compact) cube C contained in U and let $\{y_1, \dots, y_{2^n}\}$ be the vertices of the cube. Then, for each $x \in C$, we can write $x = \sum_i \alpha_i y_i$, with $\sum_i \alpha_i = 1$. By the semi-concavity of f we have, for each $i = 1, \dots, 2^n$,

$$f(y_i) - f(x) \leq \langle l_x, y_i - x\rangle + \|y_i - x\|\omega(\|y_i - x\|);$$

multiplying by α_i and summing over i, we get

$$\sum_i \alpha_i f(y_i) \leq f(x) + \sum_i \alpha_i \|y_i - x\|\omega(\|y_i - x\|) \leq f(x) + B,$$

with $B = D_C\omega(D_C)$, where D_C is the diameter of the compact cube C. It follows that

$$\forall x \in C, \quad f(x) \geq \min_i f(y_i) - B.$$

We now know that f is locally bounded. Using this fact, it is not difficult to show (i). In fact, suppose that the closed ball $\bar{B}(x_0, 2r), r < +\infty$, is contained in U. For $x \in \bar{B}(x_0, r)$, we have $x - rv \in \bar{B}(x_0, 2r) \subset U$ for each $v \in \mathbb{R}^n$ with $\|v\| = 1$, and therefore

$$f(x - rv) - f(x) \leq \langle l_x, -rv\rangle + \|-rv\|\omega(\|-rv\|) = -r\langle l_x, v\rangle + r\omega(r).$$

Since, by the compactness of $\bar{B}(x_0, 2r)$, we already know that $\tilde{B} = \sup_{z \in \bar{B}(x_0, 2r)} |f(z)|$ is finite, this implies

$$\langle l_x, v\rangle \leq \frac{f(x) - f(x - rv)}{r} + \omega(r) \leq \frac{2\tilde{B}}{r} + \omega(r).$$

It follows that, for $x \in \bar{B}(x_0, r)$,

$$\|l_x\| \leq \frac{2\tilde{B}}{r} + \omega(r).$$

Since the compact set $K \subset U$ can be covered by a finite numbers of balls $\bar{B}(x_i, r_i)$, $i = 1, \dots, \ell$, we obtain (i).

To prove (ii), we consider a compact subset $K \subset U$, and we apply (i) to obtain the constant A. We denote by D_K the (finite) diameter of the compact set K. For each $x, y \in K$,

$$\begin{aligned} f(y) - f(x) &\le \langle l_x, y - x\rangle + \|y - x\|\omega(\|y - x\|) \\ &\le (\|l_x\| + \omega(D_K))\, \|y - x\| \\ &\le (A + \omega(D_K))\|y - x\|. \end{aligned}$$

Exchanging the role of x and y, we conclude that f is Lipschitz on K. □

Let us recall that a Lipschitz real valued function defined on an open subset of a Euclidean space is differentiable almost everywhere (with respect to the Lebesgue measure). Therefore by part (ii) of Lemma 6.1.5 above we obtain the following corollary:

Corollary 6.1.6. *A locally semi-concave real valued function defined on an open subset of a Euclidean space is differentiable almost everywhere with respect to the Lebesgue measure.*

In fact, in the case of semi-concave functions there is a better result which is given in Theorem 6.1.8 below, whose proof can be found in [41, Section 4.1]. Let us first give a definition:

Definition 6.1.7. We say that $E \subset \mathbb{R}^n$ is *countably* $(n-1)$-*Lipschitz* if there exists a countable family of compact subsets $K_j \subset \mathbb{R}^n$ such that:

1. E is contained in $\cup_j K_j$;

2. for each j there exists a hyperplane $H_j \subset \mathbb{R}^n = H_j \oplus H_j^\perp$, where $H_j^\perp$ is the Euclidean orthogonal of H_j, such that K_j is contained in the graph of a Lipschitz function $f_j : A_j \to H_j^\perp$ defined on a compact subset $A_j \subset H_j$.

Note that in the definition above, by the graph property (ii), the compact subset K_j has finite $(n-1)$-dimensional Hausdorff measure. Therefore any $(n-1)$-Lipschitz set is contained in a Borel (in fact σ-compact) $(n-1)$-Lipschitz set with σ-finite $(n-1)$-dimensional Hausdorff measure.

Theorem 6.1.8. *If $\varphi : U \to \mathbb{R}$ is a semi-concave function defined on the open subset U of $\mathbb{R}^n$, then φ is differentiable at each point in the complement of Borel countably $(n-1)$-Lipschitz set.*

In order to extend the definition of locally semi-concave to functions defined on a manifold, it suffices to show that this definition is stable by composition with diffeomorphisms.

Lemma 6.1.9. *Let $U, V \subset \mathbb{R}^n$ be open subsets. Suppose that $F : V \to U$ is a C^1 map. If $f : U \to \mathbb{R}$ is a locally semi-concave function then $f \circ F : V \to \mathbb{R}$ is also locally semi-concave. Moreover, if F is of class C^2, and $f : U \to \mathbb{R}$ is a locally semi-concave function with a linear modulus then $f \circ F : V \to \mathbb{R}$ is also locally semi-concave with a linear modulus.*

Proof. Since the nature of the result is local, without loss of generality we can assume that $f : U \to \mathbb{R}$ is semi-concave with modulus ω. We now show that, for every V' convex open subset whose closure $\bar{V}'$ is compact and contained in V, the restriction $f \circ F|_{V'} : V' \to \mathbb{R}$ is a semi-concave function. We set $C_{\bar{V}'} = \max_{z \in \bar{V}'} \|D_z F\|$, and we denote by $\hat{\omega}_{\bar{V}'}$ a modulus of continuity for the continuous function $z \mapsto D_z F$ on the compact subset $\bar{V}'$.

For each x, y in the compact convex subset $\bar{V}' \subset V$, we have

$$\begin{aligned} f(F(y)) - f(F(x)) &\leq \langle l_{F(x)}, F(y) - F(x) \rangle + \|F(y) - F(x)\| \omega(\|F(y) - F(x)\|) \\ &\leq \langle l_{F(x)}, DF(x)(y - x) \rangle + \|l_{F(x)}\| \hat{\omega}_{\bar{V}'}(\|y - x\|) \|y - x\| \\ &+ C_{\bar{V}'} \|y - x\| \omega(C_{\bar{V}'} \|y - x\|); \end{aligned}$$

Since $F(\bar{V}')$ is a compact subset of U we can apply part (i) of Lemma 6.1.5 to obtain that $\tilde{C}_{\bar{V}'} = \sup_{\bar{V}'} \|l_{F(x)}\|$ is finite. This implies that $f \circ F$ on V' is semi-concave with the modulus

$$\tilde{\omega}(r) = \tilde{C}_{\bar{V}'} \hat{\omega}_{\bar{V}'}(r) + C_{\bar{V}'} \omega(C_{\bar{V}'} r).$$

If F is C^2, then its derivative DF is locally Lipschitz on U, and we can assume that $\hat{\omega}_{\bar{V}'}$ is a linear modulus. Therefore, if ω is a linear modulus, we obtain that $\tilde{\omega}$ is also a linear modulus. □

Thanks to the previous lemma, we can define a locally semi-concave function (resp. a locally semi-concave function for a linear modulus) on a manifold as a function whose restrictions to charts is, when computed in coordinates, locally semi-concave (resp. locally semi-concave for a linear modulus). Moreover, it suffices to check this locally semi-concavity in charts for a family of charts whose domains of definition cover the manifold. It is not difficult to see that Theorem 6.1.8 is valid on any (second countable) manifold, since we can cover such a manifold by the domains of definition of a countable family of charts.

Now we want to introduce the notion of uniformly semi-concave family of functions.

Definition 6.1.10. Let $f_i : U \to \mathbb{R}$, $i \in I$, be a family of functions defined on an open subset U of $\mathbb{R}^n$. We will say that the family $(f_i)_{i\in I}$ is *uniformly ω-semi-concave*, where ω is a modulus of continuity, if each f_i is ω-semi-concave. We will say that the family $(f_i)_{i\in I}$ is *uniformly semi-concave* if there exists a modulus of continuity ω such that the family $(f_i)_{i\in I}$ is uniformly ω-semi-concave. We will say that the family $(f_i)_{i\in I}$ is *uniformly semi-concave with a linear modulus*, if it is uniformly ω-semi-concave, with ω of the form $t \mapsto kt$, where k is a fixed constant.

Theorem 6.1.11. *Suppose that $f_i : U \to \mathbb{R}$, $i \in I$, is a family of functions defined on an open subset U of $\mathbb{R}^n$. Suppose that this family $(f_i)_{i\in I}$ is uniformly ω-semi-concave, where ω is a modulus of continuity. If the function*

$$f(x) = \inf_{i\in I} f_i(x)$$

is finite everywhere on U, then $f : U \to \mathbb{R}$ is also ω-semi-concave.

Proof. Fix $x_0 \in U$. We can find a sequence i_n such that $f_{i_n}(x_0) \searrow f(x_0) > -\infty$. We choose a cube $C \subset U$ with center x_0. Call $y_1, \dots, y_{2^n}$ the vertices of C. By the argument in the beginning of the proof of Lemma 6.1.5, we have

$$\forall x \in C,\ \forall i \in I, \quad \min_{1\le j\le 2^n} f_i(y_j) \le f_i(x_0) + D_C\omega(D_C),$$

where D_C is the diameter of the compact cube C. Using the fact that $f(y_j) = \inf_{i\in I} f_i(y_j)$ is finite, it follows that there exists $A \in \mathbb{R}$ such that

$$\forall x \in C,\ \forall i \in I, \quad f_i(x) \ge A.$$

Choose now $\varepsilon > 0$ such that $\bar{B}(x_0, \varepsilon) \subset C$. If $l_i : \mathbb{R}^n \to \mathbb{R}$ is a linear form such that

$$\forall y \in U, \quad f_i(y) \le f_i(x_0) + \langle l_i, y - x_0\rangle + \|y - x_0\|\omega(\|y - x_0\|),$$

we obtain that, for every $v \in \mathbb{R}^n$ of norm 1,

$$A \le f_i(x_0) + \langle l_i, \varepsilon v\rangle + \varepsilon\omega(\varepsilon).$$

Since $f_{i_n}(x_0) \searrow f(x_0)$, we can assume $f_{i_n}(x_0) \le M < +\infty$ for all n, that implies

$$\|l_{i_n}\| \le \frac{M - A}{\varepsilon} + \omega(\varepsilon) < +\infty.$$

Up to extracting a subsequence, we can assume $l_{i_n} \to l$ in $\mathbb{R}^{n*}$, the dual space of $\mathbb{R}^n$. Then, as for every $y \in U$ we have $f(y) \le f_{i_n}(y)$, passing to the limit in n in the inequality

$$f(y) \le f_{i_n}(x_0) + \langle l_{i_n}, y - x_0\rangle + \|y - x_0\|\omega(\|y - x_0\|),$$

we get

$$f(y) \le f(x_0) + \langle l, y - x_0\rangle + \|y - x_0\|\omega(\|y - x_0\|).$$

Since $x_0 \in U$ is arbitrary, this concludes the proof. □

Before generalizing the notion of uniformly semi-concave family of functions to manifolds, let us look at the following example.

Example 6.1.12. For $k \in \mathbb{R}$, define $f_k : \mathbb{R} \to \mathbb{R}$ as $f_k(x) = kx$. It is clear that the family $(f_k)_{k\in\mathbb{R}}$ is ω-semi-concave for every modulus of continuity ω. In fact

$$f_k(y) - f_k(x) = k(y-x) \leq k(y-x) + |y-x|\omega(|y-x|),$$

since $\omega \geq 0$. Consider now the diffeomorphism $\varphi : \mathbb{R}^*_+ \to \mathbb{R}^*_+$, $\varphi(x) = x^2$. Then there does not exist a non-empty open subset $U \subset \mathbb{R}^*_+$, and a modulus of continuity ω, such that the family $(f_k \circ \varphi|_U)_{k\in\mathbb{R}}$ is (uniformly) ω-semi-concave. Suppose in fact, by absurd, that

$$f_k \circ \varphi(y) - f_k \circ \varphi(x) \leq l_x(y-x) + |y-x|\omega(|y-x|),$$

where l_x depends on k but not ω. Since $f_k \circ \varphi$ is differentiable we must have $l_x(y-x) = (f_k \circ \varphi)'(x)(y-x) = 2kx(y-x)$. Therefore we should have

$$ky^2 - kx^2 \leq 2kx(y-x) + |y-x|\omega(|y-x|).$$

Fix $x, y \in U$, with $y \neq x$ and set $h = y - x$. Then

$$kh^2 \leq |h|\omega(|h|) \Rightarrow k \leq \frac{\omega(|h|)}{|h|} \quad \forall k,$$

that is obviously absurd.

Therefore the following is the only reasonable definition for the notion of a uniformly locally semi-concave family of functions on a manifold.

Definition 6.1.13. We will say that the family of functions $f_i : M \to \mathbb{R}$, $i \in I$, defined on the manifold M, is *uniformly locally semi-concave* (resp. *with a linear modulus*), if we can find a cover $(U_j)_{j\in J}$ of M by open subsets, with each U_j domain of a chart $\varphi_j : U_j \xrightarrow{\sim} V_j \subset \mathbb{R}^n$ (where n is the dimension of M), such that for every $j \in J$ the family of functions $(f_i \circ \varphi_j^{-1})_{i\in I}$ is a uniformly semi-concave family of functions on the open subset V_j of $\mathbb{R}^n$ (resp. with a linear modulus).

The following corollary is an obvious consequence of Theorem 6.1.11.

Corollary 6.1.14. *If the family $f_i : M \to \mathbb{R}$, $i \in I$ is uniformly locally semi-concave (resp. with a linear modulus) and the function*

$$f(x) = \inf_{i\in I} f_i(x)$$

is finite everywhere, then $f : M \to \mathbb{R}$ is locally semi-concave (resp. with a linear modulus).

Definition 6.1.15. Suppose $c : M \times N \to \mathbb{R}$ is a function defined on the product of the manifold M by the topological space N. We will say that the family of functions $(c(\cdot, y))_{y\in N}$ is *locally uniformly locally semi-concave* (resp. *with a linear modulus)*, if for each $y_0 \in N$ we can find a neighborhood V_0 of y_0 in N such that the family $(c(\cdot, y))_{y\in V_0}$ is uniformly locally semi-concave on M (resp. with a linear modulus).

Proposition 6.1.16. *Suppose $c : M \times N \to \mathbb{R}$ is a function defined on the product of the manifold M by the topological space N, such that the family of functions $(c(\cdot, y))_{y\in N}$ is locally uniformly locally semi-concave (resp. with a linear modulus). If $K \subset N$ is compact, and the function*

$$f_K(x) = \inf_{y\in K} c(x, y)$$

is finite everywhere on U, then $f_K : U \to \mathbb{R}$ is locally semi-concave on M (resp. with a linear modulus).

Proof. By compactness of K, we can find a finite family $V_i, i = 1, \dots, \ell$ of open subsets of N such that $K \subset \cup_{i=1}^{\ell} V_i$, and for every $i = 1, \dots, \ell$, the family $(c(\cdot, y))_{y\in V_i}$ is locally uniformly locally semi-concave (resp. with a linear modulus). The function

$$f_i(x) = \inf_{y\in K\cap V_i} c(x, y)$$

is finite everywhere on U, because $f_i \geq f_K$. It follows from Corollary 6.1.14 that f_i is locally semi-concave on M (resp. with a linear modulus), for $i = 1, \dots, \ell$. Since $f_K = \min_{i=1}^{\ell} f_i$, we can apply part (ii) of Proposition 6.1.4 to conclude that f_K has the same property. □

Proposition 6.1.17. *If $c : M \times N \to \mathbb{R}$ is a locally semi-concave function (resp. with a linear modulus) on the product of the manifolds M and N, then the family of functions on M $(c(\cdot, y))_{y\in N}$ is locally uniformly locally semi-concave (resp. with a linear modulus).*

Proof. We can cover $M \times N$ by a family $(U_i \times W_j)_{i\in I, j\in J}$ of open sets with U_i open in M, W_j open in N, where U_i is the domain of a chart $\varphi_i : U_i \xrightarrow{\sim} \tilde{U}_i \subset \mathbb{R}^n$ (where n is the dimension of M), and W_j is the domain of a chart $\psi_j : W_j \xrightarrow{\sim} \tilde{W}_j \subset \mathbb{R}^m$ (where m is the dimension of M), and such that

$$(\tilde{x}, \tilde{y}) \mapsto c\left(\varphi_i^{-1}(\tilde{x}), \psi_j^{-1}(\tilde{y})\right)$$

is $\omega_{i,j}$-semi-concave on $\tilde{U}_i \times \tilde{W}_j$, for some modulus $\omega_{i,j}$. It is then clear that the family

$$(c(\varphi_i^{-1}(\tilde{x}), \psi_j^{-1}(\tilde{y})))_{\tilde{y}\in\tilde{W}_j}$$

is uniformly locally $\omega_{i,j}$-semi-concave on $\tilde{U}_i$. □

The following corollary is now an obvious consequence of Propositions 6.1.17 and 6.1.16.

Corollary 6.1.18. *Suppose $c : M \times N \to \mathbb{R}$ is a locally semi-concave function (resp. with a linear modulus) on the product of the manifolds M and N. Let K be a compact subset of N. If the function*

$$f_K(x) = \inf_{y \in K} c(x, y)$$

is finite everywhere on U, then $f_K : U \to \mathbb{R}$ is locally uniformly locally semi-concave (resp. with a linear modulus).

We end this section with another useful theorem. The proof we give is an adaptation of the proof of [64, Lemma 3.8, page 494].

Theorem 6.1.19. *Let $\varphi_1, \varphi_2 : M \to \mathbb{R}$ be two functions, with φ_1 locally semi-convex (i.e. $-\varphi_1$ locally semi-concave), and φ_2 locally semi-concave. Assume that $\varphi_1 \leq \varphi_2$. If we define $\mathcal{E} = \{x \in M \mid \varphi_1(x) = \varphi_2(x)\}$, then both φ_1 and φ_2 are differentiable at each $x \in \mathcal{E}$ with $d_x\varphi_1 = d_x\varphi_2$ at such a point. Moreover, the map $x \mapsto d_x\varphi_1 = d_x\varphi_2$ is continuous on $\mathcal{E}$.*

If φ_1 is locally semi-convex and φ_2 is locally semi-concave, both with a linear modulus, then, in fact, the map $x \mapsto d_x\varphi_1 = d_x\varphi_2$ is locally Lipschitz on $\mathcal{E}$.

Proof. Since the statement is local in nature, we will assume that $M = \overset{\circ}{\mathbb{B}}$ is the Euclidean unit ball of center 0 in $\mathbb{R}^n$, and that $-\varphi_1$ and φ_2 are semi-concave with (common) modulus ω. Suppose now that $x \in \mathcal{E}$. We can find two linear maps $l_{1,x}, l_{2,x} : \mathbb{R}^n \to \mathbb{R}$ such that

$$\begin{aligned}\varphi_1(y) &\geq \varphi_1(x) + l_{1,x}(y - x) - \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}})\\ \varphi_2(y) &\leq \varphi_2(x) + l_{2,x}(y - x) + \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}).\end{aligned}$$

Using $\varphi_1 \leq \varphi_2$, and $\varphi_1(x) = \varphi_2(x)$, we obtain

$$\begin{aligned}l_{1,x}(y - x) - \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}) \leq \varphi_1(y) - \varphi_1(x) \leq\\ \leq \varphi_2(y) - \varphi_2(x) \leq l_{2,x}(y - x) + \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}).\end{aligned} \quad (6.1.1)$$

In particular, we get

$$l_{1,x}(y - x) - \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}) \leq l_{2,x}(y - x) + \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}),$$

replacing y by $x + v$ with $\|v\|_{\text{euc}}$ small, we conclude

$$l_{1,x}(v) - \|v\|_{\text{euc}}\omega(\|v\|_{\text{euc}}) \leq l_{2,x}(v) + \|v\|_{\text{euc}}\omega(\|v\|_{\text{euc}}).$$

Therefore
$$|[l_{2,x} - l_{1,x}](v)| \le 2\|v\|_{\text{euc}}\omega(\|v\|_{\text{euc}}),$$
for v small enough. Since $l_{2,x}-l_{1,x}$ is linear it must be identically 0. We set $l_x = l_{2,x} = l_{1,x}$. For $i = 1, 2$ and $y \in \overset{\circ}{\mathbb{B}}$, we obtain from (6.1.1)
$$|\varphi_i(y) - \varphi_i(x) - l_x(y - x)| \le \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}). \tag{6.1.2}$$
This implies that φ_i is differentiable at $x \in \mathcal{E}$, with $d_x\varphi_i = l$. It remains to show the continuity of the derivative. Fix $r < 1$. We now find a modulus of continuity of the derivative on the ball $r\,\overset{\circ}{\mathbb{B}}$. If $y_1, y_2 \in \mathcal{E} \cap r\,\overset{\circ}{\mathbb{B}}$, and $\|k\|_{\text{euc}} \le 1 - r$, we can apply three times (6.1.2) to obtain
$$\begin{aligned}
\varphi_1(y_2) - \varphi_1(y_1) - d_{y_1}\varphi_1(y_2 - y_1) &\le \|y_2 - y_1\|_{\text{euc}}\omega(\|y_2 - y_1\|_{\text{euc}})\\
\varphi_1(y_2 + k) - \varphi_1(y_2) - d_{y_2}\varphi_1(k) &\le \|k\|_{\text{euc}}\omega(\|k\|_{\text{euc}})\\
-\varphi_1(y_2 + k) + \varphi_1(y_1) + d_{y_1}\varphi_1(y_2 + k - y_1) &\le \|y_2 + k - y_1\|_{\text{euc}}\omega(\|y_2 + k - y_1\|_{\text{euc}}).
\end{aligned}$$
If we add the first two inequality to the third one, we obtain
$$\begin{aligned}
[d_{y_1}\varphi_1 - d_{y_2}\varphi_1](k) \le{}& \|y_2 - y_1\|_{\text{euc}}\omega(\|y_2 - y_1\|_{\text{euc}}) + \|k\|_{\text{euc}}\omega(\|k\|_{\text{euc}})\\
&+ [\|y_2 - y_1\|_{\text{euc}} + \|k\|_{\text{euc}}]\omega(\|y_2 - y_1\|_{\text{euc}} + \|k\|_{\text{euc}}),
\end{aligned}$$
which implies, exchanging k with $-k$, and using that the modulus ω is non-decreasing
$$|[d_{y_1}\varphi_1 - d_{y_2}\varphi_1](k)| \le 2[\|y_2 - y_1\|_{\text{euc}} + \|k\|_{\text{euc}}]\omega(\|y_2 - y_1\|_{\text{euc}} + \|k\|_{\text{euc}}).$$
Since $\|y_2 - y_1\|_{\text{euc}} < 2$, we can apply the inequality (1.5.3) above with any k such that $\|k\|_{\text{euc}} = (1 - r)\|y_2 - y_1\|_{\text{euc}}/2$. If we divide the inequality (1.5.3) by $\|k\|_{\text{euc}}$, and take the sup over all k such that $\|k\|_{\text{euc}} = (1 - r)\|y_2 - y_1\|_{\text{euc}}/2$, we obtain
$$\|d_{y_1}\varphi_1 - d_{y_2}\varphi_1\|_{\text{euc}} \le 2\Big[\frac{2}{1 - r} + 1\Big]\omega\big(\big(1 + \frac{1 - r}{2}\big)\|y_2 - y_1\|_{\text{euc}}\big).$$
It follows that a modulus of continuity of $x \mapsto d_x\varphi_1$ on $\mathcal{E} \cap r\,\overset{\circ}{\mathbb{B}}$ is given by
$$t \mapsto \frac{6 - 2r}{1 - r}\omega(\frac{3 - r}{2}t).$$
This implies the continuity of the map $x \mapsto d_x\varphi_1$ on $\mathcal{E} \cap r\,\overset{\circ}{\mathbb{B}}$. It also shows that it is Lipschitz on $\mathcal{E} \cap r\,\overset{\circ}{\mathbb{B}}$ when ω is a linear modulus. □

6.2 Tonelli Lagrangians

6.2.1 Definition and background

We recall some of the basic definition, and some of the results in Calculus of variations (in one variable). There are a lot of references on the subject. In [65], one can find an introduction to the subject that is particularly suited for our purpose. Other references are [38] and the first chapters in [112]. A brief and particularly nice description of the main results is contained in [45].

Definition 6.2.1 (Lagrangian). If M is a manifold, a *Lagrangian* on M is a function $L : TM \to \mathbb{R}$. In the following we will assume that L is at least bounded below and continuous.

Definition 6.2.2 (Action). If L is a Lagrangian on M, for an absolutely continuous curve $\gamma : [a, b] \to M, a \leq b$, we can define its *action* $\mathbb{A}_L(\gamma)$ by

$$\mathbb{A}_L(\gamma) = \int_a^b L(\gamma(s), \dot{\gamma}(s))\, ds.$$

Note that the integral is well defined with values in $\mathbb{R} \cup \{+\infty\}$, because L is bounded below, and $s \to L(\gamma(s), \dot{\gamma}(s))$ is defined a.e. and measurable. To make things simpler to write, we set $\mathbb{A}_L(\gamma) = +\infty$ if γ is not absolutely continuous.

Definition 6.2.3 (Minimizer). If L is a Lagrangian on the manifold M, an absolutely continuous curve $\gamma : [a, b] \to M$, with $a \leq b$, is an *L-minimizer*, if $\mathbb{A}_L(\gamma) \leq \mathbb{A}_L(\delta)$ for every absolutely continuous curve $\delta : [a, b] \to M$ with the same endpoints, i.e. such that $\delta(a) = \gamma(a)$ and $\delta(b) = \gamma(b)$.

Definition 6.2.4 (Tonelli Lagrangian). We will say that $L : TM \to \mathbb{R}$ is a *weak Tonelli Lagrangian* on M, if it satisfies the following hypotheses:

(a) L is C^1;

(b) for each $x \in M$, the map $L(x, \cdot) : T_x M \to \mathbb{R}$ is strictly convex;

(c) there exist a complete Riemannian metric g on M and a constant $C > -\infty$ such that

$$\forall (x, v) \in TM, \quad L(x, v) \geq \|v\|_x + C$$

where $\|\cdot\|_x$ is the norm on $T_x M$ obtained from the Riemannian metric g;

(d) for every compact subset $K \subset M$ the restriction of L to $T_K M = \cup_{x\in K} T_x M$ is superlinear in the fibers of $TM \to M$: this means that for every $A \geq 0$, there exists a constant $C(A, K) > -\infty$ such that

$$\forall (x, v) \in T_K M, \quad L(x, v) \geq A\|v\|_x + C(A, K).$$

We will say that L is a *Tonelli Lagrangian*, if it is a weak Tonelli Lagrangian, and satisfies the following two strengthening of conditions (a) and (b) above:

(a') L is C^2;

(b') for every $(x, v) \in TM$, the second partial derivative $\dfrac{\partial^2 L}{\partial v^2}(x, v)$ is positive definite on $T_x M$.

Since above a compact subset of a manifold all Riemannian metrics are equivalent, if condition (d) in the definition is satisfied for one particular Riemannian metric, then it is satisfied for any other Riemannian metric.

Note that when L is a weak Tonelli Lagrangian on M, and $U : M \to \mathbb{R}$ is a C^1 function *which is bounded below*, then $L + U$, defined by $(L + U)(x, v) = L(x, v) + U(x)$ is a weak Tonelli Lagrangian. If moreover L is a Tonelli Lagrangian, and U is C^2 and bounded below, then $L + U$ is a Tonelli Lagrangian. Therefore one can generate a lot of (weak) Tonelli Lagrangians from the following simple example.

Example 6.2.5. Suppose that g is a complete smooth Riemannian metric on M, and $r > 1$. We define the Lagrangian $L_{r,g}$ on M by

$$L_{r,g}(x, v) = \|v\|_x^r = g_x(v, v)^{r/2}.$$

1) $L_{2,g}$ is a Tonelli Lagrangian.

2) For any $r > 1$, the Lagrangian is C^1 and is a weak Tonelli Lagrangian.

In both cases, the Riemannian metric mentioned in condition (c) of Definition 6.2.4 is the same metric g.

Moreover, the vertical derivative of the Lagrangian $L_{r,g}$ is given by

$$\frac{\partial L_{r,g}}{\partial v}(x, v) = r\|v\|_x^{r-2} g_x(v, \cdot),$$

Proof. Since $r > 1$ it is not difficult to check that L has (in coordinates) partial derivatives everywhere with

$$\frac{\partial L_{r,g}}{\partial x}(x, 0) = 0 \quad \text{and} \quad \frac{\partial L_{r,g}}{\partial v}(x, 0) = 0,$$

and that these partial derivatives are continuous. Therefore L is C^1. A simple computation gives

$$\frac{\partial L_{r,g}}{\partial v}(x,v) = r\|v\|_x^{r-2} g_x(v,\cdot).$$

We now prove condition (c) and (d) of Definition 6.2.4 at once. In fact, if A is given, we have

$$L_{r,g}(x,v) = \|v\|_x^r \geq A\|v\|_x - A^{r/r-1},$$

as on can see by considering separately the two cases $\|v\|_x^{r-1} \geq A$ and $\|v\|_x^{r-1} \leq A$. The rest of the proof is easy. □

The completeness of the Riemannian metric in condition (c) of Definition 6.2.4 above is crucial to guarantee that a set of the form

$$\mathcal{F} = \{\gamma \in C^0([a,b],M) \mid \gamma(a) \in K,\ \mathbb{A}_L(\gamma) \leq \kappa\},$$

where K is a compact subset in M, κ is a finite constant, and $a \leq b$, is compact in the C^0 topology. In fact, condition (c) implies that the curves in such a set $\mathcal{F}$ have a g-length which is bounded independently of γ. Since K is compact (assuming M connected to simplify things) this implies that there exist $x_0 \in M$ and $R < +\infty$ such that all the curves in $\mathcal{F}$ are contained in the closed ball $\bar{B}(x_0,R) = \{y \in M \mid d(x,y) \leq R\}$, where d is the distance associated to the Riemannian metric g. But such a ball $\bar{B}(x_0,R)$ is compact since g is complete (Hopf-Rinow Theorem). From there, one obtains that the set $\mathcal{F}$ is compact in the C^0 topology, see [38, Chapters 2 and 3].

The direct method in the Calculus of Variations, see [38, Theorem 3.7, page 114] or [65] for Tonelli Lagrangians, implies:

Theorem 6.2.6. *Suppose L is a weak Tonelli Lagrangian on the connected manifold M. Then for every $a,b \in \mathbb{R}$, $a < b$, and every $x,y \in M$, there exists an absolutely continuous curve $\gamma : [a,b] \to M$ which is an L-minimizer with $\gamma(a) = x$ and $\gamma(b) = y$.*

In fact in [38, Theorem 3.7, page 114], the existence of absolutely continuous minimizers is valid under very general hypotheses on the Lagrangian L (the C^1 hypothesis on L is much stronger than necessary). We now come to the problem of regularity of minimizers which uses the C^1 hypothesis on L:

Theorem 6.2.7. *If L is a weak Tonelli Lagrangian, then every minimizer $\gamma : [a,b] \to M$ is C^1. Moreover, on every interval $[t_0,t_1]$ contained in a domain of a chart, it satisfies the following equality written in the coordinate system*

$$\frac{\partial L}{\partial v}(\gamma(t_1),\dot\gamma(t_1)) - \frac{\partial L}{\partial v}(\gamma(t_0),\dot\gamma(t_0)) = \int_{t_0}^{t_1} \frac{\partial L}{\partial x}(\gamma(s),\dot\gamma(s))\,ds, \tag{6.2.1}$$

which is an integrated from of the Euler-lagrange equation. This implies that $\partial L/\partial v(\gamma(t),\dot\gamma(t))$ is a C^1 function of t with

$$\frac{d}{dt}\left[\frac{\partial L}{\partial v}(\gamma(t),\dot\gamma(t))\right]=\frac{\partial L}{\partial x}(\gamma(t),\dot\gamma(t)).$$

Moreover, if L is a C^r Tonelli Lagrangian, with $r\geq 2$, then any minimizer is of class C^r.

Proof. We will only sketch the proof. If L is a Tonelli Lagrangian, this theorem would be a formulation of what is nowadays called Tonelli's existence and regularity theory. In that case its proof can be found in many places, for example [38], [45], or [65]. The fact that the regularity of minimizers holds for C^1 (or even less smooth) Lagrangians is more recent. The fact that a minimizer is Lipschitz has been established by Clarke and Vinter, see [44, Corollary 1, page 77, and Corollary 3.1, page 90] (again the hypothesis L is C^1 is stronger than the one required in this last work). The same fact under weaker regularity assumptions on L has been proved in [6]. A short and elegant proof of the fact that a minimizer for the class of absolutely continuous curves is necessarily Lipschitz has been given by Clarke, see [46]. Once one knows that γ is Lipschitz, when L is C^1 it is possible to differentiate the action, see [38], [45], or [65], and, using an integration by parts, one can show that γ satisfies the following integrated form of the Euler-Lagrange equation for almost every $t\in[t_0,t_1]$, for some fixed linear form c:

$$\frac{\partial L}{\partial v}(\gamma(t),\dot\gamma(t))=c+\int_{t_0}^{t}\frac{\partial L}{\partial x}(\gamma(s),\dot\gamma(s))\,ds. \tag{6.2.2}$$

But the continuity of the right hand side in (6.2.2) implies that $\partial L/\partial v(\gamma(t),\dot\gamma(t))$ extends continuously everywhere on $[t_0,t_1]$. Conditions (a) and (b) on L imply that the global Legendre transform

$$\mathscr{L}: TM\to T^*M,$$
$$(x,v)\mapsto (x,\frac{\partial L}{\partial v}(x,v)),$$

is continuous and injective, therefore a homeomorphism on its image by, for example, Brouwer's Theorem on the Invariance of Domain (see also Proposition 6.2.9 below). We therefore conclude that $\dot\gamma(t)$ has a continuous extension to $[t_0,t_1]$. Since γ is Lipschitz this implies that γ is C^1. Equation (6.2.1) follows from (6.2.2), which now holds everywhere by continuity. □

In fact we will only use the cases when L is C^2, in which case this regularity of minimizers will follow from the "usual" Tonelli regularity theory, or when L is of the form $L(x,v)=\|v\|_x^p$, $p>1$, where the norm is obtained from a C^2 Riemannian metric, in which case the minimizers are necessarily geodesics which are of course as smooth as the Riemannian metric, see Proposition 6.2.24 below.

To obtain further properties it is necessary to introduce the global Legendre transform.

Definition 6.2.8 (Global Legendre Transform). If L is a C^1 Lagrangian on the manifold L, its *global Legendre transform* $\mathscr{L} : TM \to T^*M$, where T^*M is the cotangent bundle of M, is defined by

$$\mathscr{L}(x,v) = (x, \frac{\partial L}{\partial v}(x,v)).$$

Proposition 6.2.9. *If L is a weak Tonelli Lagrangian on the manifold M, then its global Legendre transform $\mathscr{L} : TM \to T^*M$ is a homeomorphism from TM onto T^*M.*

Moreover, if L is a C^r Tonelli Lagrangian with $r \geq 2$, then $\mathscr{L}$ is C^{r-1}.

Proof. We first prove the surjectivity of $\mathscr{L}$. Suppose $p \in T_x^*M$. By condition (d) in Definition 6.2.4, we have

$$\begin{aligned} p(v) - L(x,v) &\leq p(v) - (\|p\|_x + 1)\|v\|_x - C(\|p\|_x + 1, \{x\}) \\ &\leq -\|v\|_x - C(\|p\|_x + 1, \{x\}). \end{aligned}$$

But this last quantity tends to $-\infty$, as $\|v\|_x \to +\infty$. Therefore the continuous function $v \mapsto p(v) - L(x,v)$ achieves a maximum at some point $v_p \in T_xM$. Since this function is C^1, its derivative at v_p must be 0. This yields $p - \partial L/\partial v(x, v_p) = 0$. Hence $(x,p) = \mathscr{L}(x, v_p)$.

To prove injectivity of $\mathscr{L}$, it suffices to show that for $v, v' \in T_xM$, with $v \neq v'$, we have $\partial L/\partial v(x,v) \neq \partial L/\partial v(x,v')$. Consider the function $\varphi : [0,1] \to \mathbb{R}, t \mapsto L(x, tv + (1-t)v')$, which by condition (b) of Definition 6.2.4 is strictly convex. Since it is C^1, we must have $\varphi'(0) \neq \varphi'(1)$. In fact, if that was not the case, then the non-decreasing function φ' would be constant on $[0,1]$, and φ would be affine on $[0,1]$. This contradicts strict convexity. By a simple computation, we therefore get

$$\frac{\partial L}{\partial v}(x,v')(v - v') = \varphi'(0) \neq \varphi'(1) = \frac{\partial L}{\partial v}(x,v)(v - v').$$

This implies $\partial L/\partial v(x,v') \neq \partial L/\partial v(x,v)$. We now show that $\mathscr{L}$ is a homeomorphism. Since this map is continuous, and bijective, we have to check that it is proper, i.e. inverse images under $\mathscr{L}$ of compact subsets of T^*M are (relatively) compact. For this it suffices to show that for every compact subset $K \subset M$, and every $C < +\infty$, the set

$$\{(x,v) \in T_KM \mid \|\frac{\partial L}{\partial v}(x,v)\|_x \leq C\}$$

is compact. By convexity of $v \mapsto L(x,v)$, we obtain

$$\frac{\partial L}{\partial v}(x,v)(v) \geq L(x,v) - L(x,0).$$

But $\|\partial L/\partial v(x,v)\|_x \geq \partial L/\partial v(x,v)(v/\|v\|_x)$, therefore by condition (d) of Definition 6.2.4, we conclude that

$$\forall A \geq 0, \forall (x,v) \in T_K M, \quad \|\frac{\partial L}{\partial v}(x,v)\|_x \geq A - [C(K,A)/\|v\|_x].$$

Taking $A = C+1$, we get the inclusion

$$\{(x,v) \in T_K M \mid \|\frac{\partial L}{\partial v}(x,v)\|_x \leq C\} \subset \{(x,v) \in T_K M \mid \|v\|_x \leq C(K, C+1)\},$$

and the compactness of the first set follows.

Suppose now that L is a C^r Tonelli Lagrangian with $r \geq 2$. Obviously $\mathscr{L}$ is C^{r-1}. By the inverse function theorem, to show that it is a C^{r-1} diffeomorphism, it suffices to show that the derivative is invertible at each point of TM. But a simple computation in coordinates show that the derivative of $\mathscr{L}$ at (x,v) is given in matrix form by

$$\begin{pmatrix} \mathrm{Id} & 0 \\ \frac{\partial^2 L}{\partial x \partial v}(x,v) & \frac{\partial^2 L}{\partial v^2}(x,v) \end{pmatrix}$$

This is clearly invertible by (b') of Definition 6.2.4. □

Definition 6.2.10. If L is a Lagrangian on M, we define its *Hamiltonian* $H : T^*M \to \mathbb{R} \cup \{+\infty\}$ by

$$H(x,p) = \sup_{v \in T_x M} p(v) - L(x,v).$$

Proposition 6.2.11. *Let L be a weak Tonelli Lagrangian on the manifold M. Its Hamiltonian H is everywhere finite valued and satisfies the following properties:*

(a*) *H is C^1, and in coordinates*

$$\begin{cases} \frac{\partial H}{\partial p}(\mathscr{L}(x,v)) = v \\ \frac{\partial H}{\partial x}(\mathscr{L}(x,v)) = -\frac{\partial L}{\partial x}(x,v). \end{cases}$$

(b*) *for each $x \in M$, the map $H(x,\cdot) : T_x^*M \to \mathbb{R}$ is strictly convex;*

(d*) *for every compact subset $K \subset M$ the restriction of H to $T^*_K M = \cup_{x\in K} T^*_x M$ is superlinear in the fibers of $T^*M \to M$: this means that for every $A \geq 0$, there exists a finite constant $C^*(A, K)$ such that*

$$\forall (x,p) \in T^*_K M, \quad H(x,p) \geq A\|p\|_x + C^*(A,K).$$

In particular, the function H is a proper map, i.e. inverse images under H of compact subsets of $\mathbb{R}$ are compact.

If L is a C^r Tonelli Lagrangian with $r \geq 2$, then

(a'*) *H is C^r;*

(b'*) *for every $(x,v) \in M$, the second partial derivative $\dfrac{\partial^2 H}{\partial p^2}(x,p)$ is positive definite on $T^*_x M$.*

Proof. To show differentiability, using a chart in M, we can assume that $M = U$ is an open subset in $\mathbb{R}^m$. Moreover, since all Riemannian metrics are equivalent above compact subsets, replacing U by an open subset V with compact closure contained in U, we can assume that the norm used in (c) of Definition 6.2.4 is the constant standard Euclidean norm $\|\cdot\|_{\text{euc}}$ on the second factor of $TV = V \times \mathbb{R}^m$, that is

$$\forall x \in V,\ \forall v \in \mathbb{R}^m, \quad L(x,v) \geq A\|v\|_{\text{euc}} + C(A),$$

where $C(A)$ is a finite constant, and $\sup_{x\in V} L(x,0) \leq C < +\infty$.

We have $T^*V = V \times \mathbb{R}^{m*}$, where $\mathbb{R}^{m*}$ is the dual space of $\mathbb{R}^m$. We will denote by $\|\cdot\|_{\text{euc}}$ also the dual norm on $\mathbb{R}^{m*}$ obtained from $\|\cdot\|_{\text{euc}}$ on $\mathbb{R}^m$. We now fix $R > 0$. If $p \in \mathbb{R}^{m*}$ satisfies $\|p\|_{\text{euc}} \leq R$, we have

$$\begin{aligned} p(v) - L(x,v) &\leq \|p\|_{\text{euc}}\|v\|_{\text{euc}} - (R+1)\|v\|_{\text{euc}} - C(R+1) \\ &\leq -\|v\|_{\text{euc}} - C(R+1). \end{aligned}$$

Since $L(x,0) \leq C$ for $x \in V$, it follows that, for $\|v\|_{\text{euc}} > C - C(R+1)$,

$$p(v) - L(x,v) \leq -C \leq -L(x,0).$$

This implies

$$H(x,p) = \sup_{v\in\mathbb{R}^m} p(v) - L(x,v) = \sup_{\|v\|_{\text{euc}} \leq C - C(R+1)} p(v) - L(x,v),$$

Therefore the sup in the definition of $H(x,p)$ is attained at a point $v_{(x,p)}$ with $\|v_{(x,p)}\|_{\text{euc}} \leq C - C(R+1)$. Note that this point $v_{(x,p)}$ is unique (compare with the argument proving that the Legendre transform is surjective). In fact, at its maximum $v_{(x,p)}$, the C^1 function $v \mapsto p(v) - L(x,v)$ must have 0 derivative, and therefore

$$p = \frac{\partial L}{\partial v}(x, v_{(x,p)}).$$

This means $(x,p) = \mathscr{L}(x, v_{(x,p)})$, but the Legendre transform is injective by Proposition 6.2.9.

Note, furthermore, that the map

$$f : \big(V \times \{\|p\|_{\mathrm{euc}} \le R\}\big) \times \{\|v\|_{\mathrm{euc}} \le C - C(R+1)\} \to \mathbb{R},$$

$$((x,p),v) \mapsto p(v) - L(x,v),$$

is C^1. Therefore we obtain that H is C^1 from the following classical lemma whose proof is left to the reader.

Lemma 6.2.12. *Let $f : N \times K \to \mathbb{R}, (x,k) \mapsto f(x,k)$ be a continuous map, where N is a manifold, and K is a compact space. Define $F : N \to \mathbb{R}$ by $F(x) = \sup_{k \in K} f(x,k)$. Suppose that:*

1. $\dfrac{\partial f}{\partial x}(x,k)$ *exists everywhere and is continuous as a function of both variables (x,k);*
2. *for every $x \in N$, the set $\{k \in K \mid f(x,k) = F(x)\}$ is reduced to a single point, which we will denote by k_x.*

Then F is C^1, and the derivative D_xF of F at x is given by

$$D_xF = \frac{\partial f}{\partial x}(x, k_x).$$

Returning to the proof of Proposition 6.2.11, by the last statement of the above lemma we also obtain

$$\frac{\partial H}{\partial p}(x,p) = v_{(x,p)} \quad \text{and} \quad \frac{\partial H}{\partial x}(x,p) = -\frac{\partial L}{\partial x}(x, v_{(x,p)})$$

Since $(x,p) = \mathscr{L}(x, v_{(x,p)})$, this can be rewritten as

$$\frac{\partial H}{\partial p} \circ \mathscr{L}(x,v) = v \quad \text{and} \quad \frac{\partial H}{\partial x} \circ \mathscr{L}(x,v) = -\frac{\partial L}{\partial x}(x,v), \tag{6.2.3}$$

which proves (a*). Note that when L is a C^r Tonelli Lagrangian, by Proposition 6.2.9 the Legendre transform $\mathscr{L}$ is a C^{r-1} global diffeomorphism. From the expression of the partial derivatives above, we conclude that $\partial H/\partial p$ and $\partial H/\partial x$ are both C^{r-1}. This proves (a'*).

We now prove (b'*). Taking the derivative in v of the first equality in (6.2.3)

$$\frac{\partial H}{\partial p}\left[x, \frac{\partial L}{\partial v}(x,v)\right] = v,$$

we obtain the matrix equation

$$\frac{\partial^2 H}{\partial p^2}(\mathscr{L}(x,v)) \cdot \frac{\partial^2 L}{\partial v^2}(x,v) = \mathrm{Id}_{\mathbb{R}^m},$$

where the dot $\cdot$ represents the usual product of matrices. This means that the matrix representative of $\partial^2 H/\partial p^2(x,p)$ is the inverse of the matrix of a positive definite quadratic form, therefore $\partial^2 H/\partial p^2(x,p)$ is itself positive definite.

We prove (b*). Suppose $p_1 \neq p_2$ are both in T_x^*M. Fix $t \in (0,1)$, and set $p_3 = tp_1 + (1-t)p_2$. The covectors p_1, p_2, p_3 are all distinct. Call v_1, v_2, v_3 elements in T_xM such that $p_i = \partial L/\partial v(x, v_i)$. By injectivity of the Legendre transform, the tangent vectors v_1, v_2, v_3 are also all distinct. Moreover, for $i = 1, 2$ we have

$$H(x, p_i) = p_i(v_i) - L(x, v_i),$$

$$H(x, p_3) = p_3(v_3) - L(x, v_3) = t[p_1(v_3) - L(x, v_3)] + (1-t)[p_2(v_3) - L(x, v_3)].$$

Since the sup in the definition of $H(x,p)$ is attained at a unique point, and v_1, v_2, v_3 are all distinct, for $i = 1, 2$ we must have

$$p_i(v_3) - L(x, v_3) < p_i(v_i) - L(x, v_i) = H(x, p_i).$$

It follows that

$$H(x, tp_1 + (1-t)p_2) < tH(x, p_1) + (1-t)H(x, p_2).$$

It remains to prove (d*). Fix a compact set K in M. Since

$$H(x,p) \geq p(v) - L(x,v),$$

we obtain

$$H(x,p) \geq \sup_{\|v\|_x \leq A} p(v) + \inf_{x \in K, \|v\|_x \leq A} -L(x,v).$$

But $\sup_{\|v\|_x \leq A} p(v) = A\|p\|_x$, and $C^*(A,K) = \inf_{x \in K, \|v\|_x \leq A} -L(x,v)$ is finite by compactness. □

Since for a weak Tonelli Lagrangian L, the Hamiltonian $H : T^*M \to \mathbb{R}$ is C^1, we can define the Hamiltonian vector field X_H on T^*M. This is rather standard and uses the fact that the exterior derivative of the Liouville form on M defines a symplectic form on M, see [2] or [84]. The vector field X_H is entirely characterized by the fact that in coordinates obtained from a chart in M, it is given by

$$X_H(x,p) = (\frac{\partial H}{\partial p}(x,p), -\frac{\partial H}{\partial x}(x,p)).$$

So the associated ODE is given by

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial p}(x,p) \\ \dot{p} = -\dfrac{\partial H}{\partial x}(x,p). \end{cases}$$

In this form, it is an easy exercise to check that H is constant on any solution of X_H.

We know come to the simple and important connection between minimizers and solutions of X_H.

Theorem 6.2.13. *Suppose L is a weak Tonelli Lagrangian on M. If $\gamma : [a,b] \to M$ is a minimizer for L, then the Legendre transform of its speed curve $t \mapsto \mathscr{L}(\gamma(t), \dot{\gamma}(t))$ is a C^1 solution of the Hamiltonian vector field X_H obtained from the Hamiltonian H associated to L.*

Moreover, if L is a Tonelli Lagrangian, there exists a (partial) C^1 flow ϕ_t^L on TM such that every speed curve of an L-minimizer is a part of an orbit of ϕ_t^L. This flow is called the Euler-Lagrange flow, is defined by

$$\phi_t^L = \mathscr{L}^{-1} \circ \phi_t^H \circ \mathscr{L},$$

where ϕ_t^H is the partial flow of the C^1 vector filed X_H.

Proof. If we write $(x(t), p(t)) = \mathscr{L}(\gamma(t), \dot{\gamma}(t))$ then

$$x(t) = \gamma(t) \quad \text{and} \quad p(t) = \frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t)).$$

By Theorem 6.2.7, $x(t) = \gamma(t)$ is C^1 with $\dot{x}(t) = \dot{\gamma}(t)$. The fact that $p(t)$ is C^1 follows again from Theorem 6.2.7, which also yields in local coordinates

$$\dot{p}(t) = \frac{\partial L}{\partial x}(\gamma(t), \dot{\gamma}(t)).$$

Since $(x(t), p(t)) = \mathscr{L}(\gamma(t), \dot\gamma(t))$, we conclude from Proposition 6.2.11 that $t \mapsto (x(t), p(t))$ satisfies the ODE

$$\begin{cases} \dot x = \dfrac{\partial H}{\partial p}(x,p) \\ \dot p = -\dfrac{\partial H}{\partial x}(x,p). \end{cases}$$

Therefore the Legendre transform of the speed curve of a minimizer is a solution of the Hamiltonian vector field X_H.

If L is a Tonelli Lagrangian, by Proposition 6.2.11 the Hamiltonian H is C^2. Therefore the vector field X_H is C^1, and it defines a (partial) C^1 flow ϕ_t^H. The rest follows from what was obtained above and the fact that the Legendre transform is C^1. □

We recall the following definition

Definition 6.2.14 (Energy). If L is a C^1 Lagrangian on the manifold M, its *energy* $E : TM \to \mathbb{R}$ is defined by

$$E(x,v) = H \circ \mathscr{L}(x,v) = \frac{\partial L}{\partial v}(x,v)(v) - L(x,v).$$

Corollary 6.2.15 (Conservation of Energy). *If L is a C^1 Lagrangian on the manifold M, and $\gamma : [a,b] \to M$ is a C^1 minimizer for L, then the energy E is constant on the speed curve*

$$s \mapsto (\gamma(s), \dot\gamma(s)).$$

Proof. In fact $E(\gamma(s), \dot\gamma(s)) = H \circ \mathscr{L}(\gamma(s), \dot\gamma(s))$. But $s \mapsto \mathscr{L}(\gamma(s), \dot\gamma(s))$ is a solution of the vector field H, and the Hamiltonian H is constant on orbits of X_H. □

Proposition 6.2.16. *If L is a weak Tonelli Lagrangian on the manifold M, then for every compact subset $K \subset M$, and every $C < +\infty$, the set*

$$\{(x,v) \in T_K M \mid E(x,v) \le C\}$$

is compact, i.e. the map $E : TM \to \mathbb{R}$ is proper on every subset of the form $\pi^{-1}(K)$, where K is a compact subset of M.

Proof. Since $E = H \circ \mathscr{L}$, this follows from the fact that H is proper and $\mathscr{L}$ is a homeomorphism. □

Proposition 6.2.17. *Let L be a weak Tonelli Lagrangian on M. Suppose K is a compact subset of M, and $t > 0$. Then we can find a compact subset $\tilde K \subset M$ and a finite constant A, such that every minimizer $\gamma : [0,t] \to M$ with $\gamma(0), \gamma(t) \in K$ satisfies $\gamma([0,t]) \subset \tilde K$ and $\|\dot\gamma(s)\|_{\gamma(s)} \le A$ for every $s \in [0,t]$.*

Proof. We will use as a distance d the one coming from the complete Riemannian metric. All finite closed balls in this distance are compact (Hopf-Rinow theorem). We choose $x_0 \in K$, and R such that $K \subset B(x_0, R)$ (we could take $R = \mathrm{diam}(K)$, the diameter of K). We now pick $x, y \in K$. If $\alpha : [0,t] \to M$ is a geodesic with $\alpha(0) = x$, $\alpha(t) = y$ and whose length is $d(x, y)$ (such a geodesic exists by completeness), the inequality

$$d(x, y) \leq d(x, x_0) + d(x_0, y) \leq 2R$$

implies that $\alpha([0,t]) \subset \bar{B}(x_0, 3R)$. Moreover $\|\dot{\alpha}(s)\|_{\alpha(s)} = d(x, y)/t \leq 2R/t$ for every $s \in [0,t]$. By compactness, the Lagrangian L is bounded on the set

$$\mathscr{K} = \{(z, v) \in TM \mid z \in \bar{B}(x_0, 3R),\ \|v\|_z \leq 2R/t\}.$$

We call θ an upper bound of L on $\mathscr{K}$. Obviously the action of α on $[0,t]$ is less than $t\theta$, and therefore if $\gamma : [0,t] \to M$ is a minimizer with $\gamma(0), \gamma(t) \in K$, we get $\int_0^t L(\gamma(s), \dot{\gamma}(s))\, ds \leq t\theta$. Using condition (c) on the Lagrangian L and what we obtained above, we see that

$$Ct + \int_0^t \|\dot{\gamma}(s)\|_{\gamma(s)}\, ds \leq t\theta.$$

It follows that we can find $s_0 \in [0,t]$ such that

$$\|\dot{\gamma}(s_0)\|_{\gamma(s_0)} \leq \theta - C.$$

Moreover

$$\gamma([0,t]) \subset \bar{B}(\gamma(0), t(\theta - C)) \subset \bar{B}(x_0, R + t(\theta - C)).$$

We set $\tilde{K} = \bar{B}(x_0, R + t(\theta - C))$. If we define

$$\theta_1 = \sup\{E(z, v) \mid (z, v) \in TM,\ z \in \tilde{K},\ \|v\|_z \leq \theta - C\},$$

we see that θ_1 is finite by compactness. Moreover $E(\gamma(s_0), \dot{\gamma}(s_0)) \leq \theta_1$. But, as mentioned earlier, the energy $E(\gamma(s), \dot{\gamma}(s))$ is constant on the curve. This implies that the speed curve

$$s \mapsto (\gamma(s), \dot{\gamma}(s))$$

is contained in the compact set

$$\tilde{\mathscr{K}} = \{(z, v) \in TM \mid z \in \tilde{K},\ E(z, v) \leq \theta_1\}.$$

Observing that the set $\tilde{\mathscr{K}}$ does not depend on γ, this finishes the proof. □

6.2.2 Lagrangian costs and semi-concavity

Definition 6.2.18 (Costs for a Lagrangian). Suppose $L : TM \to \mathbb{R}$ is a Lagrangian on the connected manifold M, which is bounded from below. For $t > 0$, we define the *cost* $c_{t,L} : M \times M \to \mathbb{R}$ by

$$c_{t,L}(x, y) = \inf_{\gamma(0)=x,\gamma(t)=y} \mathbb{A}_L(\gamma)$$

where the infimum is taken over all the absolutely continuous curves $\gamma : [0, t] \to M$, with $\gamma(0) = x$, and $\gamma(t) = y$, and $\mathbb{A}_L(\gamma)$ is the action $\int_0^t L(\gamma(s), \dot\gamma(s))\, ds$ of γ.

Using a change of variable in the integral defining the action, it is not difficult to see that $c_{t,L} = c_{1,L^t}$ where the Lagrangian L^t on M is defined by $L^t(x, v) = tL(x, t^{-1}v)$. Observe that L^t is a (weak) Tonelli Lagrangian if L is.

Theorem 6.2.19. *Suppose that $L : TM \to \mathbb{R}$ is a weak Tonelli Lagrangian. Then, for every $t > 0$, the cost $c_{t,L}$ is locally semi-concave on $M \times M$. Moreover, if the derivative of L is locally Lipschitz, then $c_{t,L}$ is locally semi-concave with a linear modulus.*

In particular, if L is a Tonelli Lagrangian for every $t > 0$, the cost $c_{t,L}$ is locally semi-concave on $M \times M$ with a linear modulus.

Proof. By the remark preceding the statement of the theorem, it suffices to prove this for $c = c_{1,L}$. Let n be the dimension of M. Choose two charts $\varphi_i : U_i \xrightarrow{\sim} \mathbb{R}^n$, $i = 0, 1$, on M. We will show that

$$(\tilde x_0, \tilde x_1) \mapsto c(\varphi_0^{-1}(\tilde x_0), \varphi_1^{-1}(\tilde x_1))$$

is semi-concave on $\mathring{\mathbb{B}} \times \mathring{\mathbb{B}}$, where $\mathbb{B}$ is the closed Euclidean unit ball of center 0 in $\mathbb{R}^n$. By Proposition 6.2.17, we can find a constant A such that for every minimizer $\gamma : [0, 1] \to M$, with $\gamma(i) \in \varphi_i^{-1}(\mathbb{B})$, we have

$$\forall s \in [0, 1], \quad \|\dot\gamma(s)\|_{\gamma(s)} \le A.$$

We now pick $\delta > 0$ such that for all $z_1, z_2 \in \mathbb{R}^n$, with $\|z_1\|_{\text{euc}} \le 1$, $\|z_2\|_{\text{euc}} = 2$,

$$d(\varphi_i^{-1}(z_1), \varphi_i^{-1}(z_2)) \ge \delta, \quad i = 0, 1,$$

where $\|\cdot\|_{\text{euc}}$ denote the Euclidean norm. Then we choose $\varepsilon > 0$ such that $A\varepsilon < \delta$. It follows that

$$\gamma([0, \varepsilon]) \subset \varphi_0^{-1}\big(2\,\mathring{\mathbb{B}}\big) \text{ and } \gamma([1 - \varepsilon, 1]) \subset \varphi_1^{-1}\big(2\,\mathring{\mathbb{B}}\big).$$

We set $\tilde{x}_i = \varphi_i(\gamma(i))$, $i = 0, 1$. For $h_0, h_1 \in \mathbb{R}^n$ we can define $\tilde{\gamma}_{h_0} : [0, \varepsilon] \to \mathbb{R}^n$ and $\tilde{\gamma}_{h_1} : [1-\varepsilon, 1] \to \mathbb{R}^n$ as

$$\tilde{\gamma}_{h_0}(s) = \frac{\varepsilon - s}{\varepsilon} h_0 + \varphi_0(\gamma(s)), \quad 0 \leq s \leq \varepsilon,$$

$$\tilde{\gamma}_{h_1}(s) = \frac{s - (1-\varepsilon)}{\varepsilon} h_1 + \varphi_1(\gamma(s)), \quad 1 - \varepsilon \leq s \leq 1.$$

We observe that when $h_0 = 0$ (or $h_1 = 0$) the curve coincide with γ. Moreover $\tilde{\gamma}_{h_0}(0) = \tilde{x}_0 + h_0$, $\tilde{\gamma}_{h_1}(1) = \tilde{x}_1 + h_1$. We suppose that $\|h_i\|_{\text{euc}} \leq 2$. In that case the images of both $\tilde{\gamma}_{h_0}$ and $\tilde{\gamma}_{h_0}$ are contained in $4 \overset{\circ}{\mathbb{B}}$ and

$$\|\dot{\tilde{\gamma}}_{h_i}(s)\|_{\text{euc}} \leq \|h_i\|_{\text{euc}} + \|(\varphi_i \circ \gamma)'(s)\|_{\text{euc}} \leq 2 + \|(\varphi_i \circ \gamma)'(s)\|_{\text{euc}}.$$

Since we know that the speed of γ is bounded in M, we can find a constant A_1 such that

$$\forall s \in [0, \varepsilon], \quad \|\dot{\tilde{\gamma}}_{h_0}(s)\|_{\text{euc}} \leq A_1,$$

$$\forall s \in [1-\varepsilon, 1], \quad \|\dot{\tilde{\gamma}}_{h_1}(s)\|_{\text{euc}} \leq A_1.$$

To simplify a little bit the notation, we define the Lagrangian $L_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$L_i(z, v) = L(\varphi_i^{-1}(z), D[\varphi_i^{-1}](v)).$$

If we concatenate the three curves $\varphi_0^{-1} \circ \tilde{\gamma}_{h_0}$, $\gamma|_{[\varepsilon, 1-\varepsilon]}$ and $\varphi_1^{-1} \circ \tilde{\gamma}_{h_1}$, we obtain a curve in M between $\varphi_0^{-1}(\tilde{x}_0 + h_0)$ and $\varphi_1^{-1}(\tilde{x}_1 + h_1)$, and therefore

$$\begin{aligned} c\left(\varphi_0^{-1}(\tilde{x}_0 + h_0), \varphi_1^{-1}(\tilde{x}_1 + h_1)\right) \leq & \int_0^\varepsilon L_0(\tilde{\gamma}_{h_0}(t), \dot{\tilde{\gamma}}_{h_0}(t))\, dt \\ & + \int_\varepsilon^{1-\varepsilon} L(\gamma(t), \dot{\gamma}(t))\, dt + \int_{1-\varepsilon}^1 L_1(\tilde{\gamma}_{h_1}(t), \dot{\tilde{\gamma}}_{h_1}(t))\, dt. \end{aligned}$$

Hence

$$\begin{aligned} & c\left(\varphi_0^{-1}(\tilde{x}_0 + h_0), \varphi_1^{-1}(\tilde{x}_1 + h_1)\right) - c\left(\varphi_0^{-1}(\tilde{x}_0), \varphi_1^{-1}(\tilde{x}_1)\right) \\ & \qquad \leq \int_0^\varepsilon \left[L_0(\tilde{\gamma}_{h_0}(t), \dot{\tilde{\gamma}}_{h_0}(t)) - L_0(\varphi_0 \circ \gamma(t), (\varphi_0 \circ \gamma)'(t))\right] dt \\ & \qquad\qquad + \int_{1-\varepsilon}^1 \left[L_1(\tilde{\gamma}_{h_1}(t), \dot{\tilde{\gamma}}_{h_1}(t)) - L_1(\varphi_1 \circ \gamma(t), (\varphi_1 \circ \gamma)'(t))\right] dt. \end{aligned}$$

We now call ω a common modulus of continuity for the derivative DL_0 and DL_1 on the compact set $\bar{B}(0, 4) \times \bar{B}(0, A_1)$. Here DL_0 and DL_1 denote the total derivatives of

L_0 and L_1, i.e. with respect to all variables. When L has a derivative which is locally Lipschitz, then DL_0 and DL_1 are also locally Lipschitz on $\mathbb{R}^n \times \mathbb{R}^n$, and the modulus ω can be taken linear. Since $\tilde\gamma_{h_i}(s) \in \overset{\circ}{B}(0,4)$ and $\|\dot{\tilde\gamma}_{h_i}(s)\| \le A_1$, we get the estimate

$$\begin{aligned} c\big(\varphi_0^{-1}(\tilde x_0 + h_0), &\varphi_1^{-1}(\tilde x_1 + h_1)\big) - c\big(\varphi_0^{-1}(\tilde x_0), \varphi_1^{-1}(\tilde x_1)\big) \\ &\le \int_0^{\varepsilon} DL_0\left(\varphi_0\circ\gamma(t), (\varphi_0\circ\gamma)'(t)\right)\left(\frac{\varepsilon - t}{\varepsilon}h_0, -\frac{1}{\varepsilon}h_0\right) dt \\ &+ \int_{1-\varepsilon}^{1} DL_1\left(\varphi_1\circ\gamma(t), (\varphi_1\circ\gamma)'(t)\right)\left(\frac{t-(1-\varepsilon)}{\varepsilon}h_1, \frac{1}{\varepsilon}h_1\right) dt \\ &+ \omega\left(\frac{1}{\varepsilon}\|h_0\|_{\text{euc}}\right)\frac{1}{\varepsilon}\|h_0\|_{\text{euc}} + \omega\left(\frac{1}{\varepsilon}\|h_1\|_{\text{euc}}\right)\frac{1}{\varepsilon}\|h_1\|_{\text{euc}}. \end{aligned}$$

We observe that the sum of the first two terms in the right hand side is linear, while the sum of the last two is bounded by

$$\frac{1}{\varepsilon}\omega\left(\frac{1}{\varepsilon}\|(h_0,h_1)\|_{\text{euc}}\right)\|(h_0,h_1)\|_{\text{euc}}.$$

Therefore we obtain that

$$(\tilde x_0, \tilde x_1) \mapsto c\left(\varphi_0^{-1}(\tilde x_0), \varphi_1^{-1}(\tilde x_1)\right)$$

is semi-concave for the modulus $\tilde\omega(r) = \frac{1}{\varepsilon}\omega\left(\frac{1}{\varepsilon}r\right)$ on $\overset{\circ}{\mathbb{B}} \times \overset{\circ}{\mathbb{B}}$, as wanted. □

Corollary 6.2.20. *If L is a weak Tonelli Lagrangian on the connected manifold M, then, for every $t > 0$, a superdifferential of $c_{t,L}(x,y)$ at (x_0,y_0) is given by*

$$(v,w) \mapsto \frac{\partial L}{\partial v}(\gamma(t),\dot\gamma(t))(w) - \frac{\partial L}{\partial v}(\gamma(0),\dot\gamma(0))(v),$$

where $\gamma : [0,t] \to M$ is a minimizer for L with $\gamma(0) = x_0$, $\gamma(t) = y_0$, and $(v,w) \in T_xM \times T_yM = T_{(x,y)}(M \times M)$.

Proof. Again we will do it only for $t = 1$. If we use the notation introduced in the previous proof, we see that a superdifferential of

$$(\tilde x_0, \tilde x_1) \mapsto c\left(\varphi_0^{-1}(\tilde x_0), \varphi_1^{-1}(\tilde x_1)\right)$$

is given by

$$(h_0,h_1) \mapsto l_0(h_0) + l_1(h_1),$$

where

$$l_0(h_0) = -\int_0^{\varepsilon} \Big[\frac{t-\varepsilon}{\varepsilon}\frac{\partial L_0}{\partial x}\left(\varphi_0\circ\gamma(t),(\varphi_0\circ\gamma)'(t)\right)(h_0) \\ + \frac{1}{\varepsilon}\frac{\partial L_0}{\partial v}\left(\varphi_0\circ\gamma(t),(\varphi_0\circ\gamma)'(t)\right)(h_0)\Big]\,dt,$$

$$l_1(h_1) = \int_{1-\varepsilon}^{1} \Big[\frac{t-(1-\varepsilon)}{\varepsilon}\frac{\partial L_1}{\partial x}\left(\varphi_1\circ\gamma(t),(\varphi_1\circ\gamma)'(t)\right)(h_1) \\ + \frac{1}{\varepsilon}\frac{\partial L_1}{\partial v}\left(\varphi_1\circ\gamma(t),(\varphi_1\circ\gamma)'(t)\right)(h_1)\Big]\,dt.$$

By Theorem 6.2.7, the curve $t \mapsto \varphi_0\circ\gamma(t)$ is a C^1 extremal of L_0 and it satisfies the following integrated form of the Euler-Lagrange equation:

$$\frac{\partial L_0}{\partial v}\left(\varphi_0\circ\gamma(t),(\varphi_0\circ\gamma)'(t)\right) - \frac{\partial L_0}{\partial v}\left(\varphi_0\circ\gamma(0),(\varphi_0\circ\gamma)'(0)\right) \\ = \int_0^t \frac{\partial L_0}{\partial x}\left(\varphi_0\circ\gamma(s),(\varphi_0\circ\gamma)'(s)\right)\,ds.$$

This gives us

$$l_0(h_0) = -\frac{\partial L_0}{\partial v}\left(\varphi_0\circ\gamma(0),(\varphi_0\circ\gamma)'(0)\right) \\ -\frac{1}{\varepsilon}\int_0^{\varepsilon}\frac{d}{ds}\left[(t-\varepsilon)\int_0^t \frac{\partial L_0}{\partial x}\left(\varphi_0\circ\gamma(s),(\varphi_0\circ\gamma)'(s)\right)\,ds\right]\,dt.$$

Obviously the second term in the right hand side is 0 and so l_0 reinterpreted on $T_{x_0}M$ rather than on $\mathbb{R}^n$ gives $-\frac{\partial L}{\partial v}(\gamma(0),\dot\gamma(0))$. The treatment for l_1 is the same. □

We have avoided the first variation formula in the proof of Corollary 6.2.20, because this is usually proven for C^2 variation of curves and C^2 Lagrangians. Of course, our argument to prove this Corollary is basically a proof for the first variation formula for C^1 Lagrangians. This is of course already known and the proof is the standard one.

6.2.3 The twist condition for costs obtained from Lagrangians

Lemma 6.2.21. *Let L be a weak Tonelli Lagrangian on the connected manifold M. Suppose that L satisfies the following condition:*

(UC) *If* $\gamma_i : [a_i, b_i] \to M, i = 1, 2$ *are two* L*-minimizers such that* $\gamma_1(t_0) = \gamma_2(t_0)$ *and* $\dot\gamma_1(t_0) = \dot\gamma_2(t_0)$*, for some* $t_0 \in [a_1, b_1] \cap [a_2, b_2]$*, then* $\gamma_1 = \gamma_2$ *on the whole interval* $[a_1, b_1] \cap [a_2, b_2]$*.*

Then, for every $t > 0$*, the cost* $c_{t,L} : M \times M \to \mathbb{R}$ *satisfies the left (and the right) twist condition of Definition 1.2.4.*

Moreover, if $(x, y) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$*, then we have:*

(i) *there is a unique* L*-minimizer* $\gamma : [0, t] \to M$ *such that* $x = \gamma(0)$*, and* $y = \gamma(t)$*;*

(ii) *the speed* $\dot\gamma(0)$ *is uniquely determined by the equality*

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot\gamma(0)).$$

Proof. We first prove part (ii). Pick $\gamma : [0, t] \to M$ an L-minimizer with $x = \gamma(0)$ and $y = \gamma(t)$. From Corollary 6.2.20 we obtain the equality

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot\gamma(0)). \tag{6.2.4}$$

Since the C^1 map $v \mapsto L(x, v)$ is strictly convex, the Legendre transform $v \in T_xM \mapsto \partial L/\partial v(x, v)$ is injective, and therefore $\dot\gamma(0) \in T_xM$ is indeed uniquely determined by Equation (6.2.4) above. This proves (ii).

To prove statement (i), consider another L-minimizer $\gamma_1 : [0, t] \to M$ is $x = \gamma_1(0)$. By what we just said, we also have

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot\gamma_1(0)).$$

By the uniqueness already proved in statement (ii), we get $\dot\gamma_1(0) = \dot\gamma(0)$. It now follows from condition (UC) that $\gamma = \gamma_1$ on the whole interval $[0, t]$.

The twist condition follows easily. Consider $(x, y), (x, y_1) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$ such that

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = \frac{\partial c_{t,L}}{\partial x}(x, y_1). \tag{6.2.5}$$

By (i) there is a unique L-minimizer $\gamma : [0, t] \to M$ (resp. $\gamma_1 : [0, t] \to M$) such that $x = \gamma(0), y = \gamma(1)$ (resp. $x = \gamma_1(0), y_1 = \gamma_1(1)$), and

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot\gamma(0)) \quad \text{and} \quad \frac{\partial c_{t,L}}{\partial x}(x, y_1) = -\frac{\partial L}{\partial v}(x, \dot\gamma_1(0)).$$

From equation (6.2.5), and the injectivity of the Legendre transform of L, it follows that $\dot\gamma_1(0) = \dot\gamma(0)$. From condition (UC) we get $\gamma = \gamma_1$ on the whole interval $[0, t]$. In particular, we obtain $y = \gamma(t) = \gamma_1(t) = y_1$. □

The next lemma is an easy consequence of Lemma 6.2.21 above.

Lemma 6.2.22. *Let L be a weak Tonelli Lagrangian on M. If we can find a continuous local flow ϕ_t defined on TM such that:*

(UC') *for every L-minimizer $\gamma : [a,b] \to M$, and every $t_1, t_2 \in [a,b]$, the point $\phi_{t_2-t_1}(\gamma(t_1), \dot\gamma(t_1))$ is defined and $(\gamma(t_2), \dot\gamma(t_2)) = \phi_{t_2-t_1}(\gamma(t_1), \dot\gamma(t_1))$,*

then L satisfies (UC)*. Therefore, for every $t > 0$, the cost $c_{t,L} : M \times M \to \mathbb{R}$ satisfies the left twist (and the right) condition of Definition 1.2.4.*

Moreover, if $(x,y) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$, then $y = \pi\phi_t(x,v)$, where $\pi : TM \to M$ is the canonical projection, and $v \in T_xM$ is uniquely determined by the equation

$$\frac{\partial c_{t,L_{r,g}}}{\partial x}(x,y) = -\frac{\partial L}{\partial v}(x,v).$$

The curve $s \in [0,t] \mapsto \pi\phi_s(x,v)$ is the unique L-minimizer $\gamma : [0,t] \to M$ with $\gamma(0) = x, \gamma(1) = y$.

Note that the following proposition is contained in Theorem 6.2.13.

Proposition 6.2.23. *If L is a Tonelli Lagrangian, then it satisfies condition* (UC') *for the Euler Lagrange flow ϕ^L_t.*

Proposition 6.2.24. *Suppose g is a complete Riemannian metric on the connected manifold M, and $r > 1$. For a given $t > 0$, the cost $c_{t,L_{r,g}}$ of the weak Tonelli Lagrangian $L_{r,g}$, defined by*

$$L_{r,g}(x,v) = \|v\|^r_x = g_x(v,v)^{r/2},$$

is given by

$$c_{t,L_{r,g}} = t^{r-1} d^r_g(x,y),$$

where d_g is the distance defined by the Riemannian metric. The Lagrangian $L_{r,g}$ satisfies condition (UC') *of Lemma 6.2.22 for the geodesic flow ϕ^g_t of g. Therefore its cost $c_{t,L_{r,g}}$ satisfies the left (and the right) twist condition. Moreover, if $(x,y) \in \mathcal{D}(\Lambda^l_{c_{t,L_{r,g}}})$, then $y = \pi\phi^g_t(x,v)$, where $\pi : TM \to M$ is the canonical projection, and $v \in T_xM$ is uniquely determined by the equation*

$$\frac{\partial c_{t,L_{r,g}}}{\partial x}(x,y) = -\frac{\partial L_{r,g}}{\partial v}(x,v).$$

Proof. Define s by $1/s + 1/r = 1$. Let $\gamma : [a,b] \to M$ be a piecewise C^1 curve. Denoting by $\ell_g(\gamma)$ the Riemannian length of γ, we can apply Hölder inequality to obtain

$$\int_a^b \|\gamma(s)\|_x \, ds \le (b-a)^{1/s} \left(\int_a^b \|\gamma(s)\|^r_x \, ds\right)^{1/r},$$

with equality if and only if γ is parameterized with $\|\gamma(s)\|_x$ constant, i.e. proportionally to arc-length. This of course implies

$$(b-a)^{-r/s}\ell_g(\gamma)^r \leq \int_a^b \|\gamma(s)\|_x^r \, ds,$$

with equality if and only if γ is parameterized proportionally to arc-length. Since any curve can be reparametrized proportionally to arc-length and $r/s = r-1$, we conclude that

$$c_{t,L_{r,g}}(x,y) = t^{1-r} d_g(x,y)^r,$$

and that an $L_{r,g}$-minimizing curve has to minimize the length between its end-points. Therefore any $L_{r,g}$-minimizing curve is a geodesic and its speed curve is an orbit of the geodesic flow ϕ_t^g. Therefore $L_{r,g}$ satisfies condition (UC') of Lemma 6.2.22 for the geodesic flow ϕ_t^g of g. The rest of the proposition follows from Lemma 6.2.22. □

Bibliography

[1] A.Abbondandolo & A.Figalli: *High action orbits for Tonelli Lagrangians and superlinear Hamiltonians on compact configuration spaces.* J. Differential Equations, **234** (2007), no.2, 626-653.

[2] R.Abraham & J.E.Marsden: *Foundations of mechanics. Second edition, revised and enlarged.* (1978) Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass.

[3] L.Ambrosio: *Lecture notes on optimal transport problems,* in Mathematical Aspects of Evolving Interfaces, Lecture Notes in Math. **1812**, Springer-Verlag, Berlin/New York (2003), 1-52.

[4] L.Ambrosio: *Transport equation and Cauchy problem for BV vector fields.* Invent. Math., **158** (2004), no.2, 227-260.

[5] L.Ambrosio: *Lecture notes on transport equation and Cauchy problem for non-smooth vector fields.* Preprint, 2005 (available at *http://cvgmt.sns.it/people/ambrosio*).

[6] L.Ambrosio, O.Ascenzi & G.Buttazzo: *Lipschitz regularity for minimizers of integral functionals with highly discontinuous integrands.* J. Math. Anal. Appl., **142** (1989), no. 2, 301-316.

[7] L.Ambrosio & G.Crippa: *Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields.* UMI-Lecture Notes, to appear.

[8] L.Ambrosio & A.Figalli: *Geodesics in the space of measure-preserving maps and plans.* Arch. Ration. Mech. Anal., to appear.

[9] L.Ambrosio & A.Figalli: *On the regularity of the pressure field of Brenier's weak solutions to incompressible Euler equations.* Calc. Var. Partial Differential Equations, **31** (2007), no. 4, 497-509.

[10] L.Ambrosio, N.Fusco & D.Pallara: *Functions of bounded variation and free discontinuity problems.* Oxford Mathematical Monographs, 2000.

[11] L.Ambrosio, N.Gigli & G.Savaré: *Gradient flows in metric spaces and in the Wasserstein space of probability measures.* Lectures in Mathematics, ETH Zurich, Birkhäuser (2005).

[12] L.Ambrosio, B.Kirchheim & A.Pratelli: *Existence of optimal transport maps for crystalline norms.* Duke Math. J., **125** (2004), no. 2, 207-241.

[13] L.Ambrosio & A.Pratelli: *Existence and stability results in the L^1 theory of optimal transportation.* Lectures notes in Mathematics, **1813**, Springer Verlag, Berlin/New York (2003), 123-160.

[14] L.Ambrosio, S.Lisini & G.Savaré: *Stability of flows associated to gradient vector fields and convergence of iterated transport maps.* Manuscripta Math., **121** (2006), 1-50.

[15] V.Arnold: *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits.* (French) Ann. Inst. Fourier (Grenoble), **16** (1966), fasc. 1, 319-361.

[16] V.Bangert: *Analytische Eigenschaften konvexer Funktionen auf Riemannschen Manigfaltigkeiten.* J. Reine Angew. Math., **307** (1979), 309-324.

[17] V.Bangert: *Minimal measures and minimizing closed normal one-currents.* Geom. Funct. Anal., **9** (1999), 413-427.

[18] S.Bates: *Toward a precise smoothness hypothesis in Sard's Theorem.* Proc. Amer. Math. Soc., **117** (1993), no. 1, 279-283.

[19] J.-D. Benamou & Y. Brenier: *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.* Numer. Math., **84** (2000), 375-393.

[20] P.Bernard: *Existence of $C^{1,1}$ critical sub-solutions of the Hamilton-Jacobi equation on compact manifolds.* Annales scientifiques de l'ENS, to appear.

[21] P.Bernard: *Young measures, superposition, and transport.* In preparation.

[22] P.Bernard & B.Buffoni: *Optimal mass transportation and Mather theory.* J. Eur. Math. Soc., **9** (2007), no. 1, 85-121.

[23] P.Bernard & B.Buffoni: *The Monge problem for supercritical Mañé potential on compact manifolds.* Adv. Math., **207** (2006), no. 2, 691-706.

[24] M.BERNOT: *Irrigation and Optimal Transport*, Ph.D. Thesis, École Normale Supérieure de Cachan, 2005. Available at *http://www.umpa.ens-lyon.fr/˜mbernot.*

[25] M.BERNOT, V.CASELLES & J.M.MOREL: *Traffic plans.* Publ. Mat., **49** (2005), no. 2, 417-451.

[26] M.BERNOT, V.CASELLES & J.M.MOREL: *The structure of branched transportation networks.* Calc. Var. Partial Differential Equations, to appear.

[27] M.BERNOT & A.FIGALLI: *Synchronized traffic plans and stability of optima.* ESAIM Control Optim. Calc. Var., to appear.

[28] F.BOUCHUT *Renormalized solutions to the Vlasov equations with coefficient of bounded variations* Arch. Rational Mech. Anal., **157** (2001), 75-90.

[29] A.BRANCOLINI, G.BUTTAZZO & F.SANTAMBROGIO: *Path functionals over Wasserstein spaces.* J. Eur. Math. Soc., **8** (2006) 3, 414-434.

[30] Y.BRENIER: *Polar decomposition and increasing rearrangement of vector fields.* C. R. Acad. Sci. Paris Sér. I Math., **305** (1987), no. 19, 805-808.

[31] Y.BRENIER: *The least action principle and the related concept of generalized flows for incompressible perfect fluids.* J. Amer. Math. Soc., **2** (1989), 225-255.

[32] Y. BRENIER: *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math., **44** (1991), 375-417.

[33] Y.BRENIER: *The dual least action problem for an ideal, incompressible fluid.* Arch. Rational Mech. Anal., **122** (1993), 323-351.

[34] Y.BRENIER: *A homogenized model for vortex sheets.* Arch. Rational Mech. Anal., **138** (1997), 319-353.

[35] Y.BRENIER: *Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations.* Comm. Pure Appl. Math., **52** (1999), 411-452.

[36] Y.BRENIER: *Convergence of the Vlasov-Poisson system to the incompressible Euler equations.* Comm. Partial Differential Equations, **25** (2000), no. 3-4, 737-754.

[37] Y.BRENIER & W.GANGBO: *L^p approximation of maps by diffeomorphisms.* Calc. Var. Partial Differential Equations, **16** (2003), no. 2, 147-164.

[38] G.BUTTAZZO, M.GIAQUINTA & S. HILDEBRANDT: *One-dimensional variational problems.* Oxford Lecture Series in Mathematics and its Applications, **15** (1998), Oxford Univ. Press, New York.

[39] L.CAFFARELLI, M.FELDMAN & R.J.MCCANN: *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs.* J. Amer. Math. Soc., **15** (2002), 1-26.

[40] L.A.CAFFARELLI & S.SALSA, EDITORS: *Optimal transportation and applications.* Lecture Notes in Mathematics, **1813**. Lectures from the C.I.M.E. Summer School held in Martina Franca, September 2–8, 2001, Springer-Verlag, Berlin (2003).

[41] P.CANNARSA & C.SINESTRARI: *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control.* Progress in Nonlinear Differential Equations and Their Applications, **58**, Birkhauser Boston (2004).

[42] J.A.CARRILLO, R.J.MCCANN & C.VILLANI: *Contractions in the 2-Wasserstein length space and thermalization of granular media.* Arch. Rational Mech. Anal., **179** (2006), no. 2, 217-263

[43] C.CASTAING & M.VALADIER: *Convex Analysis and Measurable Multifunctions.* Lectures Notes in Mathematics, **580**, Spinger-Verlag, Berlin-New York, 1977.

[44] F.H.CLARKE & R.B.VINTER: *Regularity properties of solutions to the basic problem in the calculus of variations.* Trans. Amer. Math. Soc., **289** (1985), 73-98.

[45] F.H. CLARKE: *Methods of dynamic and nonsmooth optimization.* CBMS-NSF Regional Conference Series in Applied Mathematics, **57** (1989), Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[46] F.H.CLARKE: *A Lipschitz regularity theorem.* Ergodic Theory Dynam. Systems, to appear.

[47] F.COLOMBINI & N.LERNER: *Uniqueness of continuous solutions for BV vector fields.* Duke Math. J., **111** (2002), 357-384.

[48] F.COLOMBINI & N.LERNER: *Uniqueness of L^∞ solutions for a class of conormal BV vector fields.* Contemp. Math., vol. 368, 133-156. AMS Providence, RI, 2005.

[49] G.CONTRERAS, J.DELGADO & R.ITURRIAGA: *Lagrangians flows: the dynamics of globally minimizing orbits. II.* Bol. Soc. Brasil. Mat. (N.S.), **28** (1997), no. 2, 155-196.

[50] D.CORDERO-ERASQUIN, R.J.MCCANN & M.SCHMUCKENSCHLAGER: *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb.* Invent. Math., **146** (2001), no. 2, 219-257.

[51] D.L.COHN: *Measure theory.* Birkhäuser Boston, (1980), Mass.

[52] G.CONTRERAS: *Action potential and weak KAM solutions.* Calc. Var. Partial Differential Equations, **13** (2001), no. 4, 427-458.

[53] D'ARCY THOMPSON: *On Growth and Form.* Cambridge University Press, 1942.

[54] R.J.DIPERNA & P.-L.LIONS: *On the Fokker-Planck-Boltzmann equation.* Comm. Math. Phys., **120** (1988), 1-23.

[55] R.J.DIPERNA & P.-L.LIONS: *On the Cuachy problem for Boltzmann equations: global existence and weak stability.* Annals of Math., **130** (1989), no.2, 321-366.

[56] R.J.DIPERNA & P.-L.LIONS: *Ordinary differential equations, transport theory and Sobolev spaces.* Invent. Math., **98** (1989), no.3, 511-547.

[57] R.M.DUDLEY: *Real Analysis and Probability.* Cambridge University Press, 2002.

[58] D.G.EBIN & J.MARSDEN: *Groups of diffeomorphisms and the motion of an ideal incompressible fluid.* Annals of Math., **2** (1970), 102–163.

[59] L.C.EVANS: *Partial differential equations and Monge-Kantorovich mass transfer*, in R. Bott et al., editors, *Current Developments in Mathematics.* (1997), International Press, Cambridge, 26-78.

[60] L.C.EVANS & W.GANGBO: *Differential equations methods for the Monge-Kantorovich mass transfer problem.* Mem. Amer. Math. Soc., **137** (1999).

[61] L.C.EVANS & R.F.GARIEPY: *Measure Theory and Fine Properties of Functions.* Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.

[62] A.FATHI: *Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens.* C. R. Acad. Sci. Paris Sér. I Math., **324** (1997), no. 9, 1043-1046.

[63] A.FATHI: *Solutions KAM faibles conjuguées et barrières de Peierls.* C. R. Acad. Sci. Paris Sér. I Math., **325** (1997), no. 6, 649-652.

[64] A.FATHI: *Regularity of C^1 solutions of the Hamilton-Jacobi equation.* Ann. Fac. Sci. Toulouse Math. , **12** (2003), 479-516.

[65] A.Fathi: *Weak KAM theorems in Lagrangian Dynamics.* Book to appear.

[66] A.Fathi & A.Figalli: *Optimal transportation on non-compact manifolds.* Israel J. Math., to appear.

[67] A.Fathi, A.Figalli & L.Rifford: *On a problem of Mather.* Comm. Pure Appl. Math., to appear.

[68] A.Fathi & E.Maderna: *Weak KAM theorem on non compact manifolds.* Nonlinear Differential Equations Appl., to appear.

[69] A.Fathi & A.Siconolfi: *Existence of C^1 critical subsolutions of the Hamilton-Jacobi equation.* Invent. Math., **155** (2004), 363-388.

[70] H.Federer: *Geometric measure theory. Die Grundlehren der mathematischen Wissenschaften,* **153**, Springer-Verlag New York Inc., New York (1969).

[71] M.Feldman & R.J.McCann: *Monge's transport problem on a Riemannian manifold.* Trans. Amer. Math. Soc., **354** (2002), 1667-1697.

[72] S.Ferry: When ϵ-boundaries are manifolds. Fund. Math., **90** (1976), no. 3, 199-210.

[73] A.Figalli: *Trasporto ottimale su varietà non compatte.* Degree Thesis (in english), (2006) (available at *http://cvgmt.sns.it/people/figalli*).

[74] A.Figalli: *The Monge problem on non-compact manifolds.* Rend. Sem. Mat. Univ. Padova, **117** (2007), 147-166.

[75] A.Figalli: *Existence, Uniqueness, and Regularity of Optimal Transport Maps.* SIAM, Journal of Math. Anal., **39** (2007), no. 1, 126-137.

[76] A.Figalli: *A simple proof of the Morse-Sard theorem in Sobolev spaces.* Proc. Amer. Math. Soc., to appear.

[77] A.Figalli: *Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients.* J. Funct. Anal, **254** (2008), no.1, 109-153.

[78] A.Figalli & C.Villani: *Strong displacement convexity on Riemannian manifolds.* Math. Z., **257** (2007), no.2, 251-259.

[79] G.Forni & J.N.Mather: *Action minimizing orbits in Hamiltonian systems, Transition to chaos in classical and quantum mechanics, Montecatini Terme,1991.* Lecture Notes in Math., **1589** (1994), Springer,Berlin, 92-186.

[80] W.GANGBO & R.J.MCCANN: *The geometry of optimal transportation.* Acta Math., **177** (1996), 113-161.

[81] W.GANGBO: *The Monge mass transfer problem and its applications, in Monge Ampère equation: applications to geometry and optimization.* Contemp. Math., **226** (1999), Amer. Math. Soc., Providence, RI, 79-104.

[82] E.N.GILBERT: *Minimum cost communication networks.* Bell System Tech. J., **46** (1967), 2209-2227.

[83] M.HAURAY, C.LEBRIS & P.-L.LIONS: *Deux remarques sur les flots généralisées d'équations différentielles ordinaires. [Two remarks generalized flows for ordinary differential equations].* C. R. Acad. Sc. Paris, Sér. I, Math, submitted.

[84] H.HOFER & E.ZEHNDER: *Symplectic invariants and Hamiltonian dynamics.* Birkhauser Advanced Texts: Basler Lehrbücher. (1994) Birkhauser Verlag, Basel.

[85] L.V.KANTOROVICH: *On the transfer of masses.* Dokl. Akad. Nauk. SSSR, **37** (1942), 227-229.

[86] L.V.KANTOROVICH: *On a problem of Monge.* Uspekhi Mat.Nauk., **3** (1948), 225-226.

[87] M.KNOTT & C.S.SMITH: *On the optimal mapping of distributions.* J. Optim. Theory Appl., **43** (1984), 39-49.

[88] N.V.KRYLOV & M.RÖCKNER: *Strong solutions of stochastic equations with singular time dependent drift.* Probab. Theory Related Fields, **131** (2005), no. 2, 154-196.

[89] J.M.LASRY & P.-L.LIONS: *A remark on regularization in Hilbert spaces.* Israel J. Math., **55** (1986), no. 3, 257-266.

[90] C.LEBRIS & P.-L.LIONS: *Renormalized solutions of some transport equations with partially $W^{1,1}$ velocities and applications.* Annali di matematica pura e applicata, **183** (2004), 97-130.

[91] C.LEBRIS & P.-L.LIONS: *Existence and uniqueness of solutions to Fokker-Planck type equations with irregular coefficients.* Comm. Partial Differential Equations, to appear.

[92] C.LEBRIS & P.-L.LIONS: *Generalized flows for stochastic differential equations with irregular coefficients.* In preparation.

[93] J.-L.LIONS: *Équations diffŕentielles opérationnelles et problèmes aux limites. (French)* Die Grundlehren der mathematischen Wissenschaften, 111. Springer-Verlag, Berlin, 1961.

[94] P.-L.LIONS: *Mathematical topics in fluid mechanics, Vol. I: incompressible models.* Oxford Lecture Series in Mathematics and its applications, **3**, Oxford University Press, 1996.

[95] P.-L.LIONS: *Sur les équations différentielles ordinaires et les équations de transport. [On ordinary differential equations and transport equations].* C. R. Acad. Sc. Paris, Sér. I, Math, **326** (1998), no. 7, 833-838.

[96] P.-L.LIONS, G.PAPANICOLAU & S.R.S.VARADHAN: *Homogenization of Hamilton-Jacobi equation.* Unpublished preprint, 1987.

[97] J.LOTT & C.VILLANI: *Ricci curvature via optimal transport.* Annals of Math., to appear.

[98] F.MADDALENA, S.SOLIMINI & J.M.MOREL: *A variational model of irrigation patterns.* Interfaces and Free Boundaries, **5** (2003), 391–416.

[99] R.MAÑÉ: *On the minimizing measures of Lagrangian dynamical systems.* Nonlinearity, **5** (1992), no. 3, 623-638.

[100] R.MAÑÉ: *Lagrangians flows: the dynamics of globally minimizing orbits.* Bol. Soc. Brasil. Mat. (N.S.), **28** (1997), no. 2, 141-153.

[101] N.G.MARKLEY: *On the number of recurrent orbit closures.* Proc. Amer. Math. Soc., **25** (1970), no. 2, 413-416.

[102] J.N. MATHER: *Existence of quasiperiodic orbits for twist homeomorphisms of the annulus.* Topology, **21** (1982), 457-467.

[103] J.N.MATHER: *Variational construction of connecting orbits.* Ann. Inst. Fourier, **43** (1993), 1349-1386.

[104] J.N.MATHER: TOTAL DISCONNECTEDNESS OF THE QUOTIENT AUBRY SET IN LOW DIMENSIONS. Comm. Pure Appl. Math., **56** (2003), 1178-1183.

[105] J.N.MATHER: *Examples of Aubry sets.* Ergod. Th. Dynam. Sys., **24** (2004), 1667-1723.

[106] MCCANN: *A convexity principle for interacting gases and equilibrium crystals.* PhD thesis, Princeton University, 1994.

[107] R.J.McCann: *Existence and uniqueness of monotone measure-preserving maps.* Duke Math. J., **80** (1995), no. 2, 309-323.

[108] R.J.McCann: *A convexity principle for interacting gases.* Adv. Math., **128** (1997), 153-179.

[109] R.J.McCann: *Polar factorization of maps on Riemannian manifolds.* Geom. Funct. Anal., **11** (2001), no. 3, 589-608.

[110] G.Monge: *Mémoire sur la Théorie des Déblais et des Remblais.* Hist. de l'Acad. des Sciences de Paris (1781), 666-704.

[111] A.P.Morse: *The behavior of a function on its critical set.* Annals of Math., **40** (1939), 62-70.

[112] J.Moser: *Selected chapters in the calculus of variations*, Lectures in Mathematics ETH Zürich, Lecture notes by Oliver Knill, Birkhäuser Verlag, Basel (2003).

[113] J.D.Murray: *Mathematical Biology.* Biomathematics texts 19, Springer, 1993.

[114] A.Norton: *A Critical Set with Nonnull Image has Large Hausdorff Dimension.* Trans. Amer. Math. Soc., **296** (1986), no. 1, 367-376.

[115] *I.Nikolaev & E.Zhuzhoma:* Flows on 2-dimensional Manifolds. Lecture Notes in Mathematics, vol 1705, Springer-Varlag, Berlin, 1999.

[116] F.Otto & C.Villani: *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality.* J. Funct. Anal., **173** (2000), no. 2, 361-400.

[117] S.T.Rachev & L. Rüschendorf: *A characterization of random variables with minimum L^2-distance.* J. Multivariate Anal., **32** (1990), 48-54. Corrigendum in J. Multivariate Anal., **34** (1990), 156.

[118] S.T.Rachev & L.Ruschendorf: *Mass transportation problems.* Vol. I: Theory, Vol. II: Applications. Probability and ita applications, Springer, 1998.

[119] L.Rifford: *On viscosity solutions of certain Hamilton-Jacobi equations: Regularity results and generalized Sard's Theorems.* Comm. Partial Differential Equations, to appear.

[120] W.Schachermayer & J.Teichmann: *Characterization of optimal transport plans for the Monge-Kantorovich-problem.* Proc. Amer. Math. Soc., to appear.

[121] A.I.Shnirelman: *The geometry of the group of diffeomorphisms and the dynamics of an ideal incompressible fluid.* (Russian) Mat. Sb. (N.S.), **128 (170)** (1985), no. 1, 82–109.

[122] A.I.Shnirelman: *Generalized fluid flows, their approximation and applications.* Geom. Funct. Anal., **4** (1994), no. 5, 586–620.

[123] S.K.Smirnov: *Decomposition of solenoidal vector charges into elementary solenoids and the structure of normal one-dimensional currents.* St. Petersburg Math. J., **5** (1994), 841–867.

[124] A.Sorrentino: *On the total disconnectedness of the quotient Aubry set.* Preprint, 2006.

[125] D.W.Stroock & S.R.S.Varadhan: *Multidimensional diffusion processes.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 233. Springer-Verlag, Berlin-New York, 1979.

[126] Sturm, K.-T. *On the geometry of metric measure spaces. I.* Acta Math., **196**, 1 (2006), 65-131.

[127] Sturm, K.-T. *On the geometry of metric measure spaces. II.* Acta Math. **196**, 1 (2006), 133-177.

[128] K.T.Sturm & M.K.von Renesse: *Transport Inequalities, Gradient Estimates, Entropy and Ricci Curvature.* Comm. Pure Appl. Math., **58**, 7 (2005), 923-940.

[129] V.N.Sudakov: *Geometric problems in the theory of infinite-dimensional probability distributions.* Proc. Steklov Inst. Math., **141** (1979), 1-178.

[130] N.S.Trudinger & X.J.Wang: *On the Monge mass transfer problem.* Calc. Var. Partial Differential Equations, **13** (2001), 19-31.

[131] A.M.Turing: *The chemical basis of morphogenesis.* Phil. Trans. Soc. Lond., **B237** (1952), 37-72.

[132] C.Villani: *Topics in optimal transportation.* Graduate Studies in Mathematics, **58** (2003), American Mathematical Society, Providence, RI.

[133] C.Villani: *Optimal transport, old and new.* Lecture notes, 2005 Saint-Flour summer school, available online at *http://www.umpa.ens-lyon.fr/˜cvillani.*

[134] R.Vinter: *Optimal control.* Systems & Control: Foundations & Applications, (2000), Birkhäuser Boston Inc., Boston, MA.

[135] Q.XIA: *Optimal paths related to transport problems.* Commun. Contemp. Math., **5** (2003), no.2, 251-279.