Hamilton–Jacobi equations for controlled gradient flows
The comparison principle

Conforti, G.; Kraaij, R. C.; Tonon, D.

DOI
10.1016/j.jfa.2023.109853

Publication date
2023

Document Version
Final published version

Published in
Journal of Functional Analysis

Journal of Functional Analysis 284 (2023) 109853

Contents lists available at ScienceDirect

Journal of Functional Analysis

journal homepage: www.elsevier.com/locate/jfa

Full Length Article

Hamilton–Jacobi equations for controlled gradient flows: The comparison principle

G. Conforti [a,*,1], R.C. Kraaij [b], D. Tonon [c]

[a] *CMAP, Ecole Polytechnique, Route de Saclay, 91128, Palaiseau Cedex, France*
[b] *Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, the Netherlands*
[c] *Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova, via Trieste 63, 35121 Padova, Italy*

ARTICLE INFO

Article history:
Received 11 February 2022
Accepted 24 December 2022
Available online 25 January 2023
Communicated by Maria Colombo

Keywords:
Hamilton Jacobi equations
EVI gradient flows
Comparison principle
Tataru distance
Optimal transport

ABSTRACT

Motivated by recent developments in the fields of large deviations for interacting particle systems and mean field control, we establish a comparison principle for the Hamilton–Jacobi equation corresponding to linearly controlled gradient flows of an energy function $\mathcal{E}$ defined on a metric space (E, d). Our analysis is based on a systematic use of the regularizing properties of gradient flows in evolutional variational inequality (EVI) formulation, that we exploit for constructing rigorous upper and lower bounds for the formal Hamiltonian at hand and, in combination with the use of the Tataru's distance, for establishing the key estimates needed to bound the difference of the Hamiltonians in the proof of the comparison principle. Our abstract results apply to a large class of examples only partially covered by the existing theory, including gradient flows on Hilbert spaces and the Wasserstein space equipped with a displacement convex energy functional $\mathcal{E}$ satisfying McCann's condition.

* Corresponding author.
E-mail addresses: giovanni.conforti@polytechnique.edu (G. Conforti), r.c.kraaij@tudelft.nl (R.C. Kraaij), daniela.tonon@unipd.it (D. Tonon).

[1] Research supported by the ANR project ANR-20-CE40-0014.

https://doi.org/10.1016/j.jfa.2023.109853
0022-1236/

Contents

1. Introduction

The study of Hamilton–Jacobi (HJ) and related equations on infinite dimensional spaces is a flourishing research field. Such equations arise naturally in a great number of situations, including but certainly not limited to mean–field (or McKean–Vlasov) control problems, mean–field games and large deviation theory. This article is concerned with a specific class of infinite dimensional Hamilton–Jacobi equations having a common geometric structure that is typically encountered in the study of abstract versions of the so called Schrödinger problem (see [46,32,31,38,42] for some motivating examples) and in connection with large deviations theory [29]. At the formal level, given a metric space (E,d) where the metric d is generated by a Riemannian metric $\langle\cdot,\cdot\rangle$, the equation writes as

$$f-\lambda Hf=h, \quad Hf:=-\langle \operatorname{grad} f, \operatorname{grad}\mathcal{E}\rangle+\frac{1}{2}\|\operatorname{grad} f\|^2 \tag{1.1}$$

where grad is the gradient associated with $\langle\cdot,\cdot\rangle$. A fundamental example where equation (1.1) arises naturally in applications is that of the Wasserstein space $(E,d)=(\mathcal{P}_2(\mathbb{R}^d),W_2(\cdot,\cdot))$ equipped with an energy functional $\mathcal{E}$ satisfying McCann's condition. In this case, the underlying formal Riemannian metric is the so called Otto metric [44]. Equation (1.1) is expected to characterize the value function of the control problem

$$\sup\left\{\int_0^{+\infty} e^{-\lambda^{-1}t}[\lambda^{-1}h(\rho^u(t)) - \frac{1}{2}\|u(t)\|^2]\mathrm{d}t : \dot\rho^u = -\operatorname{grad}\mathcal{E}(\rho^u) + u,\ \rho^u(0)=\rho_0\right\}, \tag{1.2}$$

which can be interpreted as the problem of steering the gradient flow

$$\dot\rho = -\operatorname{grad}\mathcal{E}(\rho)$$

in such a way that an optimal balance is struck between the cost of controlling, modeled through the term $-\frac{1}{2}\|u(t)\|^2$, and the reward obtained, modeled by the term $\lambda^{-1}h(\rho^u(t))$. The above control problem can be written in the equivalent form

$$\sup\left\{\int_0^{+\infty} \lambda^{-1}e^{-\lambda^{-1}t}\left[h(\rho^u(t)) - \int_0^t \frac{1}{2}\|u(s)\|^2\mathrm{d}s\right]\mathrm{d}t \,\middle|\, \dot\rho^u = -\operatorname{grad}\mathcal{E}(\rho^u) + u,\ \rho^u(0)=\rho_0\right\},$$

that gains a natural interpretation in relation to the corresponding semigroup. In this manuscript we prove a comparison principle for viscosity solutions of (1.1) that holds under mild assumptions, the most relevant one being the existence of a gradient flow for the energy functional $\mathcal{E}$ in Evolutional Variational Inequality (EVI) formulation, see (EVI_κ) below. Since in most examples of interest one cannot make sense of $\operatorname{grad}\mathcal{E}$ and the Riemannian metric cannot be rigorously constructed, following [47,48,21,27,28,1,36,30] we argue, using (EVI_κ), that the Hamilton-Jacobi equation (1.1) can be replaced by two equations in terms of two operators $H_\dagger$ and $H_\ddagger$ that serve as upper and lower bounds for the formal Hamiltonian in (1.1). We then state a comparison principle in terms of the upper and lower bounds $H_\dagger$ and $H_\ddagger$ (see Definition 2.11). Following [47,48,21,27] the test functions in the domains of $H_\dagger$ and $H_\ddagger$, contain, next to the squared metric, the non-regular Tataru distance. This distance is not easy to handle when proving the existence of viscosity solutions, nevertheless the comparison principle we state is already of large interest. A refinement of the comparison principle presented here, that will be helpful for the existence of solutions, and the existence of solution itself will be published in subsequent articles. We also present some meaningful examples of applications of our main result in particular to controlled gradient flows in the Wasserstein space. Further applications to controlled gradient flows in Riemannian manifolds and Hilbert spaces are also discussed.

Hamilton–Jacobi equations in infinite dimensional spaces The theory of viscosity solutions for Hamilton–Jacobi equations on infinite dimensional spaces was initiated by Crandall and Lions in a series of papers [16–21] in the setting of Hilbert spaces or Banach spaces possessing the Radon-Nikodym property. Recent applications in large deviations [29], functional inequalities [39], statistical mechanics [5,6], and McKean-Vlasov control

[13] have motivated the development of a theory of viscosity solutions for Hamilton–Jacobi equations on metric spaces that are not necessarily Hilbert, and in particular over the space of probability measures endowed with a transport–like distance. A first approach to Hamilton–Jacobi equations on the space of probability measures exploits the possibility of lifting the space of probability distributions to the space of square integrable random variables in order to take advantage of the Hilbertian structure of the latter: we refer to [4,45,8,22] for some results recently obtained following this method. A second approach is more intrinsic and consists of working directly at the level of the space of probability measures and developing all the relevant notions therein. One can perform this using either the linear derivative, as shown in [9] in the context of McKean-Vlasov control for jump processes, or relying on the notion of subdifferential provided by optimal transport [3]. The connections between the intrinsic approach and the extrinsic notion of derivative obtained through the above mentioned lifting procedure have been clarified in [37]. In this manuscript, we follow the intrinsic approach and in particular we build on the achievements of the research program carried out by Feng and his coauthors [29,28,30], who developed a technique to deal with equations whose geometric structure is the same as (1.1) in terms of upper and lower bounds. We combine this intrinsic approach with the use of the Tataru distance function, as a penalization function in Ekeland's variational principle. Such idea has been introduced in [47,48] and then further refined in [21,27]. To the best of our knowledge, in this work we provide the first systematic implementation of Tataru's method in metric spaces that are not Hilbert: as a result, we can remove compactness assumptions on the sublevel sets of the energy $\mathcal{E}$ and for metric balls. While postponing to the next paragraph a more accurate comparison of our results with the above mentioned works, we stress that several other important contributions [1,35–37,50] adopt the intrinsic approach to show well–posedness of Hamilton–Jacobi equations on metric spaces. In all these works it is assumed that the variations of the Hamiltonian w.r.t. the measure argument can be locally controlled by the metric d. Since we require very little from the energy functional $\mathcal{E}$ beyond the existence of an EVI gradient flow, this assumption is systematically violated in most of the instances of (1.1) that we consider. This happens already in the basic example when $\mathcal{E}$ is the relative entropy and (E, d) is the Wasserstein space. It is worth mentioning that operating the formal change of variable $\tilde{f} = f - \mathcal{E}$ and setting $\lambda = 1$ allows to rewrite formally (1.1) in the form

$$f(\pi) - \frac{1}{2} \|\mathrm{grad}\, f(\pi)\|^2 + \mathcal{F}(\pi) = 0, \tag{1.3}$$

$\mathcal{F}(\pi) = \frac{1}{2} \|\mathrm{grad}\, \mathcal{E}(\pi)\|^2 + \mathcal{E}(\pi) - h(\pi)$. This equation has been often studied in the literature on infinite dimensional Hamilton–Jacobi equations. However, our main geometrical assumption, that is formally equivalent to the semiconvexity of $\mathcal{E}$, does not give the control on the growth of $\frac{1}{2} \|\mathrm{grad}\, \mathcal{E}(\pi)\|^2$ needed to successfully apply the techniques developed in the above mentioned references [1,35–37,50].

Master equation and mean field games The recent years have witnessed fundamental advances on the understanding of the master equation arising in the theory of Mean Field Games, see [11] and the recent works [50,34,33,12] for a sample of the recent progresses. Such equation aims at characterizing the limiting behavior of Nash equilibria in the many players regime and it has been noticed [7] that the master equation shares some properties with infinite dimensional Hamilton–Jacobi equations, and in particular with those characterizing the value function of McKean–Vlasov control problems. However, these two equations remain conceptually different as explained in [14]. For example, despite some analogies between the "monotonicity" assumption that is typically imposed on the coefficients of the master equation and the geodesic convexity of the energy functional $\mathcal{E}$ that underlies all our computations, these two geometrical assumptions are not directly related and enter the coefficients of the respective equations in a different way. In the recent article [34], the authors get past the classical monotonicity assumption and indeed obtain well posedness for the master equation by means of displacement convexity. Still, the equation considered there and (1.1) have a different nature.

Contribution of this work Our methods are largely inspired by ideas and techniques put forward in [47,48,21,29,28,27,25,41,30], where comparison principles for (1.1) have been proven in different contexts. Apart from [25,41], in which exploiting a Riemannian viewpoint they prove comparison principles in finite dimensional setting, we summarize here the contributions of the other papers in treating infinite–dimensional versions of (1.1).

- The works [47,48,21,27] deal with quadratic Hamiltonians on Hilbert spaces where the drift is not given by a gradient flow, but rather by a maximally dissipative operator C. (The subgradient of a proper lower semi-continuous convex functional is maximally dissipative, connecting the two equations.) We formally have

$$Hf(\pi) := \langle \operatorname{grad} f(\pi), C\pi \rangle + \frac{1}{2} \|\operatorname{grad} f(\pi)\|^2 \tag{1.4}$$

  Due to the non-compactness of the space, optimizers in the proof of the comparison principle are produced using Ekeland's variational principle. As the drift term arises from a (non-continuous) operator, the standard Hilbertian norm is not suitable to be used as a penalization function in Ekelands principle. Thus, a new metric-like object is introduced that is constructed from the norm in combination with the flow generated by C. A second innovation in this collection of papers concerns how to deal with C in giving rigorous understanding to the Hamiltonian in (1.4). Working for test functions of the type $f(\pi) = \frac{a}{2} \|\pi - \mu\|^2$, the drift term equals

$$\langle \pi - \mu, C\pi \rangle \tag{1.5}$$

which is ill-defined if π is not in the domain of C. However, using the dissipativity of C, this term can be upper bounded by

$$\langle \pi - \mu, C\mu \rangle \tag{1.6}$$

which is well defined as long as $\mu \in \mathcal{D}(C)$. This leads to a candidate upper bound for H, as long as one restricts to test-functions of metric type with $\mu \in \mathcal{D}(C)$. A similar procedure can be carried out to obtain a lower bound. Working with test functions with restrictions on μ makes it necessary to replace the standard duplication of variables by a quadruplication, where the two new variables take their values in $\mathcal{D}(C)$. The inclusion in $\mathcal{D}(C)$ for these two new variables is enforced by the addition of two new penalization terms. This procedure is to some extent analogous to the procedure that, in finite dimensional cases, forces the variables to take their values in a compact set.

- Building upon the works above, [28] introduces a more intrinsic point of view replacing C by the gradient of some energy functional $\mathcal{E}$. In [28] this is carried out for an energy functional $\mathcal{E}$ with compact sub-levelsets on a metric space. The inclusion in the domain of the gradient of $\mathcal{E}$ is now achieved by penalization with $\mathcal{E}$, whereas in the papers above, considered in the context that $C = -\operatorname{grad}\mathcal{E}$, the penalization can be interpreted as the square root of a Fisher information. This geometric interpretation effectively leads to much cleaner estimates. A second notable difference to the papers above is that the quadruplication is replaced by a duplication of variables. This comes at the cost of working with less-regular test functions in the domain of the Hamiltonians. To obtain existence of solutions, one typically starts out with more regular test-functions. It was shown in e.g. the three examples of Section 13 of [29] that for well-posedness one can connect the regular and non-regular Hamiltonians by performing an inf- and sup-convolutions on sub- and supersolutions respectively. This is reminiscent of the techniques used in the proof of the comparison principle for second order equations on finite dimensional spaces, see e.g. [15], and implies that the full procedure to obtain the comparison principle can be seen as one that involves a quadruplication just like in the papers mentioned above. In the example of Section 13.3 of [29], studying the controlled heat flow in the Wasserstein space, it is observed that the upper and lower bound that in (1.5) and (1.6) were given by the use of the dissipativity of the operator are now replaced by the use of an inequality that we recognize in our more general context as the evolutional variational inequality.
- In [30], the authors study the controlled Carleman equation. In this context the Hamiltonian is associated to the gradient flow of the entropy on the space of probability measures considered as a subset of an inverse Sobolev space. In this paper, a combination of the ideas above has been put to work, the procedure that involves a quadruplication, as above, in the sense of a standard duplication in combination with sup- and inf-convolutions, uses compactness of the sublevelsets of the energy. Also

in this work, an inequality is used that we recognize as the evolutional variational inequality with contractivity constant $\kappa = 0$.

In view of the above works, we combine their strengths and assemble the key idea's in a single unifying framework:

- We work with a geodesic metric space, where $\mathcal{E}$ and d do not necessarily have compact sublevel sets. In fact, we will allow $\mathcal{E}$ that are unbounded from below.
- We replace the variational inequalities used in the papers above by the systematic use of the evolutional variational inequality (EVI_κ). This inequality is the generalization of the one used in [30] and in a Hilbertian context implies the dissipativity of the operator C. Interpreting the variational inequalities used in the literature in the context of EVI, they correspond to the evolutional variational inequality with contractivity constant $\kappa = 0$. We will allow for negative κ also.
- We generalize the Tataru distance from Hilbert to general metric spaces and modify the distance to allow its application to gradient flows satisfying EVI with a negative contractivity constant κ.
- Instead of establishing the comparison principle via the duplication of variables combined with sup- and inf-convolutions, we perform the quadruplication of variables in a single go and introduce an argument generalizing the classical Lemma 3.1 of [15].

To summarize: the key innovation in our proof strategy is the systematic use of the properties of EVI gradient flows, in particular of their regularizing properties that include energy dissipation and distance contraction estimates. Indeed, gradient flows play a crucial role in: *a)* Defining suitable upper and lower bounds for the formal Hamiltonian that depend on $\mathcal{E}$ and d only; *b)* the construction of the Tataru distance and *c)* developing all the necessary estimates for the proof of the main result, in particular to bound the difference of the Hamiltonians in the proof of the comparison principle (see e.g. Lemma 3.3 and Lemma 3.4). Apart from our key assumptions on the properties of the geodesic metric space E and the existence of a gradient flow satisfying the evolution variational inequality, which can be considered to be standard in the literature, we assume in Assumption 2.9 that the directional derivative of the energy functional along 'regularized geodesics' can be controlled by the local slope of the energy. Thanks to the rather soft assumptions needed for our main results to apply, we are able to cover natural situations that, to the best of our knowledge and understanding, fall out of the range of existing techniques. Leaving all precise statements to section 5 below, we would like to mention that one novelty is that we can treat the case of the Wasserstein space equipped with a Rény entropy as energy functional: in this setting the underlying gradient flow is the porous medium equation [44]. Even if we restrict to the more classical setting where the energy functional is the sum of the Boltzmann entropy, a potential energy and an interaction energy, existing results (see e.g. [29,28]) require the confining potential to grow superquadratically at infinity in order to be applied, and several further restrictions are imposed on the interaction potential. Here, we allow for much more flexibility on both potentials. It is also plausible that the class of distances introduced in [26] leads to Hamilton-Jacobi equations whose uniqueness can be established by means of Theorem 2.14 though we leave it to future work to validate this conjecture, as well as to enlarge the range of applications of the comparison principle proven in this paper.

Organization The article has the following structure: in Section 2 we state our hypothesis and then proceed to the presentation of our main results. In section 3 we prove Theorem 2.14, that is the comparison principle for the upper and lower bounds $H_\dagger$ and $H_\ddagger$. Section 5 is devoted to examples of applications, whereas Section 4 reports on the fundamental properties of EVI gradient flows and the Tataru distance. Appendix A contains some background material on Ekeland's principle and Hamilton Jacobi equations.

Frequently used notation

- $B_R(\rho)$ the ball of radius R centered at ρ;
- $\bar{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$;
- $USC(E), LSC(E)$: space of upper semi-continuous and lower semi-continuous functions over E;
- $C(E)$ continuous and bounded functions over E;

2. The comparison principle

Our aim is to establish a comparison principle for viscosity solutions of equations of the form

$$f(\pi) - \lambda H f(\pi) = h(\pi), \quad \pi \in E \tag{2.1}$$

where (E, d) is a complete metric space, $\lambda > 0$ is a constant, h a real function on E and the action of the formal Hamiltonian H is given by

$$Hf(\pi) = -\langle \operatorname{grad}_\pi f(\pi), \operatorname{grad}_\pi \mathcal{E}(\pi)\rangle + \frac{1}{2} \|\operatorname{grad}_\pi f(\pi)\|^2, \tag{2.2}$$

where $\mathcal{E} : E \to (-\infty, +\infty]$ is some energy functional and gradients are taken w.r.t. a formal Riemannian structure on E. Various issues arise with the definition of H due to the presence of grad_π. Indeed a precise notion of gradient for E is difficult or impossible to give. For example, when (E, d) is the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2(\cdot,\cdot))$, in typical situations of interest, $\mathcal{E}$ is worth $+\infty$ on a dense set and nowhere differentiable, even though the subdifferential is well defined and non empty on a subset of the domain of $\mathcal{E}$. The lack of differentiability of entropic functionals is a well known issue in the theory of gradient flows and has led to the development of notions of gradient flows that do not appeal to $\operatorname{grad}_\pi \mathcal{E}$ directly: we refer to [3] for a comprehensive overview. In a certain sense, we adopt a similar strategy: instead of working with H directly, we construct suitable upper and lower bounds $H_\dagger$ and $H_\ddagger$, that depend on $\mathcal{E}$ rather than its gradient and that are tight enough for the comparison principle to hold. To construct the upper and lower bounds we partially rely on ideas put forward in [28,30] and draw inspiration from the EVI formulation of gradient flows which allows to put the considerations made therein on some important examples into a considerably more general framework. For example, an important with these work is that here we do not assume that the level sets of $\mathcal{E}$ are compact. Let us now proceed to introduce the most important concepts needed to properly define $H_\dagger$ and $H_\ddagger$.

2.1. EVI-gradient flows and statement of the main hypotheses

We work on a complete metric space (E, d) on which an extended functional $\mathcal{E} : E \to (-\infty, +\infty]$ is defined. In the sequel, we shall refer to $\mathcal{E}$ as to the energy, or entropy depending on the context. The next definition is that of local slope given in the first chapter of [3].

Definition 2.1. Let $\phi : E \to (-\infty, +\infty]$ be an extended functional with proper effective domain, i.e. $\mathcal{D}(\phi) := \{\pi \in E : \phi(\pi) < +\infty\} \neq \emptyset$. Then the local slope of ϕ at $\rho \in \mathcal{D}(\phi)$ is defined as

$$|\partial\phi|(\rho) := \begin{cases} \limsup_{\pi\to\rho} \frac{(\phi(\rho)-\phi(\pi))^+}{d(\rho,\pi)}, & \text{if } \phi(\rho) < +\infty. \\ +\infty, & \text{otherwise.} \end{cases}$$

Next, we define geodesic spaces.

Definition 2.2. (E, d) is a geodesic space, if for any $\rho, \pi \in E$ there exists a curve $(\boldsymbol{\zeta}^{\rho\to\pi}(t))_{t\in[0,1]}$ such that $\boldsymbol{\zeta}^{\rho\to\pi}(0) = \rho, \boldsymbol{\zeta}^{\rho\to\pi}(1) = \pi$ and for all $s, t \in [0, 1]$

$$d(\boldsymbol{\zeta}^{\rho\to\pi}(s), \boldsymbol{\zeta}^{\rho\to\pi}(t)) = |t - s| d(\rho, \pi). \tag{2.3}$$

Such a curve will be called *geodesic*.

Assumption 2.3 *(Metric and energy).* We make the following assumptions of the complete metric space (E, d) and the energy functional $\mathcal{E}$.

(a) (E, d) is a geodesic space.
(b) We assume that the energy functional $\mathcal{E} : E \to (-\infty, +\infty]$ is an extended functional such that:
  - It has a proper effective domain, i.e. $\mathcal{D}(\mathcal{E}) := \{\pi \in E : \mathcal{E}(\pi) < +\infty\} \neq \emptyset$.
  - It is lower semi-continuous.

Our second main assumption is the existence of an EVI gradient flow of $\mathcal{E}$. The EVI (Evolutional Variational Inequality) formulation is the strongest formulation of gradient flows in metric spaces, we refer to the monograph [3] and the more recent article [43] for an extensive study of this notion.

Definition 2.4. Given $\kappa \in \mathbb{R}$, we define *solution of the* EVI_k *inequality* a continuous curve $\gamma : [0, +\infty) \to E$ such that $\gamma((0, +\infty)) \subseteq \mathcal{D}(\mathcal{E})$ and

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\left(d^2(\gamma(t), \rho)\right) \leq \mathcal{E}(\rho) - \mathcal{E}(\gamma(t)) - \frac{\kappa}{2} d^2(\gamma(t), \rho), \quad \forall \rho \in \mathcal{D}(\mathcal{E}), t \in [0, +\infty). \quad (EVI_\kappa)$$

Here $\frac{\mathrm{d}^+}{\mathrm{d}t}$ denotes the upper right time derivative. An EVI_k *gradient flow* of $\mathcal{E}$ defined in $D \subset \overline{\mathcal{D}(\mathcal{E})}$ is a family of continuous maps $S(t) : D \to D, t \geq 0$ such that for every $\pi \in D$:

- The semigroup property holds

$$S[\pi](0) = \pi, \quad S[\pi](t+s) = S[S[\pi](t)](s) \quad \forall t, s \geq 0. \tag{2.4}$$

- The curve $(S[\pi](t))_{t \geq 0}$ is a solution to (EVI_κ).

We shall refer to $(S[\pi](t))_{t \geq 0}$ as the *gradient flow* of $\mathcal{E}$ started at π. To lighten the notation, from now on, we will denote with $(\pi(t))_{t \geq 0}$ the gradient flow $(S[\pi](t))_{t \geq 0}$.

Assumption 2.5. [Gradient flow and EVI] We assume the existence of an (EVI_κ) gradient flow of $\mathcal{E}$ defined on $D = E$.

Remark 2.6. Note that the above assumption implies that $\overline{\mathcal{D}(\mathcal{E})} = E$.

(EVI_κ) is known to have several important consequences (see [43]), including uniqueness of the gradient flow. Some of these facts, gathered at Lemma 4.1, play a crucial role in the proofs of our main results.

Remark 2.7. Note that the Hamiltonian is formally equivalent to

$$Hf(\pi) = \frac{\mathrm{d}^+}{\mathrm{d}t}\left(f(\pi(t))\right)|_{t=0} + \frac{1}{2}|\partial f|^2(\pi), \tag{2.5}$$

for $f : E \to (-\infty, +\infty)$ and $\pi \in E$. This representation is an important guideline for the construction of the lower and upper bounds.

For later use, we define the information functional as the squared slope of the energy.

Definition 2.8. We define the *information functional* $I : E \to [0, +\infty]$ as

$$I(\pi) := \begin{cases} |\partial\mathcal{E}|^2(\pi) & \pi \in \mathcal{D}(\mathcal{E}) \\ +\infty & \text{otherwise} \end{cases}.$$

The information functional is closely related to the gradient flow via the energy identity

$$\mathcal{E}(\pi(t)) - \mathcal{E}(\pi(0)) = -\int_0^t I(\pi(s))\mathrm{d}s,$$

see Lemma 4.1 for a rigorous version of the above relation. Our final condition is of non-standard nature. We assume that any geodesic can be approximated as well as needed with a smoother curve, typically but not necessarily another geodesic, along which the variations of $\mathcal{E}$ can be controlled with the slope. This last requirement is coherent with the interpretation of the metric slope as the norm of the gradient of $\mathcal{E}$. Note that, in most examples of interest, (2.7) below fails to be true if we replace $\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t)$ with an arbitrary geodesic and that in the infinite dimensional context this assumption is considerably weaker than the existence of directional derivatives of $\mathcal{E}$ along arbitrary geodesics.

Assumption 2.9. For any $\rho, \pi \in E$ satisfying $I(\rho) + \mathcal{E}(\pi) < +\infty$, there exist a geodesic $\boldsymbol{\zeta}^{\rho\to\pi}$ such that, for any $\theta > 0$, there exists $\tau > 0$ and a curve, not necessarily a geodesic, $(\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t))_{t\in[0,\tau]}$, satisfying

$$\limsup_{t\downarrow 0} \frac{d(\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t), \boldsymbol{\zeta}^{\rho\to\pi}(t))}{t} \le \theta, \tag{2.6}$$

and

$$\liminf_{t\downarrow 0} \frac{\mathcal{E}(\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t)) - \mathcal{E}(\rho)}{t} \le |\partial\mathcal{E}|(\rho)(d(\rho,\pi) + \theta). \tag{2.7}$$

Note that (2.6) implies that $\boldsymbol{\zeta}_\theta^{\rho\to\pi}(0) = \rho$.

We refer to (2.6) as to the *angle condition*. (2.7) can be interpreted as controllability of directional derivatives of regularized geodesics by the local slope.

2.2. A first attempt at defining upper and lower bounds

In light of the previous discussion, we can start developing a correct formulation of the Hamilton-Jacobi equation. In classical proofs of the comparison principle for first order Hamilton–Jacobi equations one needs to apply the Hamiltonian to distance–like test functions. In the following lines, ignoring all the technical issues, we shall derive a formal upper bound for $\pi \mapsto Hd^2(\cdot,\rho)(\pi)$ arguing on the basis of (EVI_κ) and on the following (formal) property of the distance

$$\forall \pi, \rho \in E \quad \left|\partial\left(\frac{1}{2}d^2(\cdot,\rho)\right)\right|^2(\pi) = d^2(\pi,\rho), \tag{2.8}$$

where $\left|\partial\left(\frac{1}{2}d^2(\cdot,\rho)\right)\right|(\pi)$ is the slope of the function $\frac{1}{2}d^2(\cdot,\rho)$ evaluated at π. Note that the above equation holds in the case of a smooth Riemaniann manifold. Let us now consider a test function $f^\dagger : E \to \mathbb{R}$ that is given in terms of the squared distance as $f^\dagger(\pi) = \frac{1}{2}ad^2(\pi,\rho)$ for some $\rho \in E$ and $a > 0$. Applying formally the representation of H from (2.5) and using the property (2.8) (as if $\pi \in \mathcal{D}(\mathcal{E})$), we obtain that

$$Hf^\dagger(\pi) = \frac{1}{2}a\frac{\mathrm{d}^+}{\mathrm{d}t}\left(d^2(\pi(t),\rho)\right)\Big|_{t=0} + \frac{1}{2}a^2d^2(\pi,\rho).$$

Then, applying (formally) Assumption 2.5 and being $a > 0$, we get

$$Hf^\dagger(\pi) \leq a\left[\mathcal{E}(\rho) - \mathcal{E}(\pi)\right] - a\frac{\kappa}{2}d^2(\pi,\rho) + \frac{1}{2}a^2d^2(\pi,\rho).$$

Let us note that this upper bound is proper as soon as $\mathcal{E}(\rho) < +\infty$, so that the right hand side is well defined, even though it may take the value $-\infty$. Therefore, we are led to a candidate definition for a first upper bound $H_{\mathrm{can},\dagger}$: its domain is

$$\mathcal{D}(H_{\mathrm{can},\dagger}) := \left\{ f^\dagger : E \to \mathbb{R},\ f^\dagger(\pi) = \frac{1}{2}ad^2(\pi,\rho) \,\middle|\, \forall\, a > 0, \forall\, \rho \in E : \mathcal{E}(\rho) < \infty \right\}$$

and for $f^\dagger(\pi) = \frac{1}{2}ad^2(\pi,\rho)$ we define our candidate Hamiltonian via

$$H_{\mathrm{can},\dagger}f^\dagger(\pi) := a\left[\mathcal{E}(\rho) - \mathcal{E}(\pi)\right] - a\frac{\kappa}{2}d^2(\pi,\rho) + \frac{1}{2}a^2d^2(\pi,\rho).$$

Similarly, we get a formal lower bound for a test function $f^\ddagger : E \to \mathbb{R}$ defined as $f^\ddagger(\mu) = -\frac{1}{2}ad^2(\gamma,\mu)$, $\gamma \in \mathcal{D}(\mathcal{E})$. Let

$$\mathcal{D}(H_{\mathrm{can},\ddagger}) := \left\{ f^\ddagger : E \to \mathbb{R},\ f^\ddagger(\mu) = -\frac{1}{2}ad^2(\gamma,\mu) \,\middle|\, a > 0,\ \gamma \in E : \mathcal{E}(\gamma) < \infty \right\}$$

be the corresponding domain then for $f^\ddagger(\mu) = -\frac{1}{2}ad^2(\gamma,\mu)$ we set

$$H_{\mathrm{can},\ddagger}f^\ddagger(\mu) = a\left[\mathcal{E}(\mu) - \mathcal{E}(\gamma)\right] + a\frac{\kappa}{2}d^2(\gamma,\mu) + \frac{1}{2}a^2d^2(\gamma,\mu).$$

Thus, instead of establishing the comparison principle for equation (2.1), we aim to show it for the upper and lower bound we found for our Hamiltonian, i.e. we would like to show that for every subsolution u (in a sense to be precised) of

$$f - \lambda H_{\mathrm{can},\dagger} f = h$$

and every supersolution v (in a sense to be precised) of

$$f - \lambda H_{\mathrm{can},\ddagger} f = h$$

we have $u \leq v$. Thanks to the formal inequalities this result would give a formal comparison principle for equation (1.1). The standard procedure to prove the comparison principle consists in using a doubling variables method. However, when doing this with our candidate Hamiltonian, we run into the known issue that optimal values are not attained, essentially because we are working in a infinite dimensional space. This issue is usually solved via Ekeland's variational principle (a version of which, the one used in this article, is Lemma A.1, in the appendix). Nevertheless, for our setting, in which the Hamiltonian contains an unbounded term, this is not enough. Indeed, once Ekeland variational principle gives us the unique optimizer, the standard procedure consists in finding good estimates for the difference of the Hamiltonians. Following [21,47,48,27], we need to apply the Ekeland variational principle with the Tataru distance as a penalization function which, in contrast with the usual distance d is Lipschitz along the gradient flow and allows for an efficient comparison of the difference between of the Hamiltonians. Let us now proceed to construct a version of the Tataru distance that is adapted to our scope.

2.3. The Tataru distance

The Tataru distance function, introduced in [47], is given in terms of the gradient flow generated by the energy functional $\mathcal{E}$ considered therein.

$$d_T(\pi, \rho) = \inf_{t \geq 0} \{t + d(\pi, \rho(t))\}, \quad \forall \pi, \rho \in E,$$

where $\rho(\cdot)$ is the gradient flow of $\mathcal{E}$ started at ρ. Note that d_T is not a metric due to a lack of symmetry. The two key properties of the above Tataru distance are that d_T is Lipschitz with respect to the metric d and that it behaves well with respect to the corresponding gradient flow

$$\frac{d_T(\pi(r), \rho) - d_T(\pi, \rho)}{r} \leq 1, \quad \forall \pi, \rho \in E,$$

for all $r \in \mathbb{R} \setminus \{0\}$. These properties are both based on the fact that the gradient flow considered there was contracting with respect to the metric. In our setting, we consider (EVI_κ) gradient flows and we allow negative values κ, i.e. a negatively curved space, and in this case the gradient flow is not anymore contracting. Thus, we have to work with an adjusted Tataru distance that takes care of all possible values of κ.

Definition 2.10. We define the Tataru distance $d_T : E \times E \to [0, +\infty)$ with respect to the metric d and energy $\mathcal{E}$ as

$$d_T(\pi, \rho) = \inf_{t \geq 0} \left\{ t + e^{\hat{\kappa} t} d(\pi, \rho(t)) \right\}, \quad \forall \pi, \rho \in E,$$

where $\hat{\kappa} = (0 \wedge \kappa) \leq 0$.

The precise statements and proofs of the main properties of Tataru distance are postponed to Section 4.2.

2.4. The comparison principle for a proper upper and lower bound

Now that we have defined the Tataru distance we are ready to introduce the upper and lower bounds for H for which we will actually establish the comparison principle. As we did before, we provide a heuristic argument to justify their definition. To do so, we begin by fixing a test function of the form

$$f^\dagger(\pi) = \frac{1}{2} a d^2(\pi, \rho) + b d_T(\pi, \mu) + c \tag{2.9}$$

for $a, b > 0$, $c \in \mathbb{R}$, and $\rho, \mu \in E$. As before, due to the presence of the term $\frac{1}{2} a d^2(\pi, \rho)$, we will need to require that $\mathcal{E}(\rho) < \infty$ in order to obtain a proper bound for the Hamiltonian. In order to bound the action of H on $f^\dagger$, we can rely again on the representation (2.5) and invoke the Lipschitzianity of d_T along the gradient flow (Lemma 4.3) that gives

$$\left| \frac{\mathrm{d}^+}{\mathrm{d}t} \left(d_T(\pi(t), \mu) \right) \big|_{t=0} \right| \leq 1.$$

Similarly, as the Tataru distance is Lipschitz with respect to d, then any gradient of d_T can be upper bounded by 1. Using these two properties and applying formally (EVI_κ) and (2.8) as we did before to define $H_{\mathrm{can},\dagger}$, we obtain that if $f^\dagger$ is as in (2.9):

$$\begin{aligned}
Hf^\dagger(\pi) =& \frac{1}{2} a \frac{\mathrm{d}^+}{\mathrm{d}t} \left(d^2(\pi(t), \rho) \right) \big|_{t=0} + b \frac{\mathrm{d}}{\mathrm{d}t} \left(d_T(\pi(t), \mu) \right) \big|_{t=0} \\
& + \frac{1}{2} \left| \partial \left(\frac{1}{2} a d^2(\cdot, \rho) + b d_T(\cdot, \mu) \right) \right|^2 (\pi) \\
\leq & a \left[\mathcal{E}(\rho) - \mathcal{E}(\pi) \right] - a \frac{\kappa}{2} d^2(\pi, \rho) + b \\
& \quad + \frac{1}{2} a^2 \left| \partial \left(\frac{1}{2} d^2(\cdot, \rho) \right) \right|^2 (\pi) + \frac{1}{2} 2ab \left| \partial \left(\frac{1}{2} d^2(\cdot, \rho) \right) \right| (\pi) \left| \partial d_T(\cdot, \mu) \right| (\pi) \\
& + \frac{1}{2} b^2 \left| \partial d_T(\cdot, \mu) \right|^2 (\pi) \\
\leq & a \left[\mathcal{E}(\rho) - \mathcal{E}(\pi) \right] - a \frac{\kappa}{2} d^2(\pi, \rho) + b
\end{aligned}$$

$$+\frac{1}{2}a^2d^2(\pi,\rho)+abd(\pi,\rho)\left|\partial d_T(\cdot,\mu)\right|(\pi)+\frac{1}{2}b^2\left|\partial d_T(\cdot,\mu)\right|^2(\pi)$$

$$\leq a\left[\mathcal{E}(\rho)-\mathcal{E}(\pi)\right]-a\frac{\kappa}{2}d^2(\pi,\rho)+b+\frac{1}{2}a^2d^2(\pi,\rho)+abd(\pi,\rho)+\frac{1}{2}b^2.$$

We can adapt this argument to test functions of the form

$$f^\ddagger(\mu):=-\frac{1}{2}ad^2(\gamma,\mu)-bd_T(\mu,\pi)+c,\quad a,b>0,c\in\mathbb{R},$$

by treating the term $\frac{1}{2}\left|\partial\left(-\frac{a}{2}d^2(\cdot,\gamma)-bd_T(\cdot,\pi)\right)\right|^2(\mu)$ in a slightly different way, namely[2]

$$\begin{aligned}
&\frac{1}{2}\left|\partial\left(-\frac{a}{2}d^2(\cdot,\gamma)-bd_T(\cdot,\pi)\right)\right|^2(\mu)\\
&\geq\frac{1}{2}\left(a\left|\partial\left(\frac{1}{2}d^2(\cdot,\gamma)\right)\right|-b|\partial d_T(\cdot,\pi)|\right)^2(\mu)\\
&=\frac{a^2}{2}d^2(\mu,\gamma)-abd(\mu,\gamma)|\partial d_T(\cdot,\pi)|(\mu)+\frac{b^2}{2}|\partial d_T(\cdot,\pi)|^2(\mu)\\
&\geq\frac{a^2}{2}d^2(\mu,\gamma)-abd(\mu,\gamma)|\partial d_T(\cdot,\pi)|(\mu)\\
&\geq\frac{a^2}{2}d^2(\mu,\gamma)-abd(\mu,\gamma)
\end{aligned}$$

We are thus led to consider the following definition, in which we prefer to underline the fact that the Hamiltonians are operators.

Definition 2.11.

1. For each $a>0,b>0,c\in\mathbb{R}$, and $\mu,\rho\in E:\,\mathcal{E}(\rho)<\infty$ let $f^\dagger=f^\dagger_{a,b,c,\mu,\rho}\in C(E)$ and $g^\dagger=g^\dagger_{a,b,c,\mu,\rho}\in USC(E)$ be given for any $\pi\in E$ by

$$f^\dagger(\pi):=\frac{1}{2}ad^2(\pi,\rho)+bd_T(\pi,\mu)+c$$
$$g^\dagger(\pi):=a\left[\mathcal{E}(\rho)-\mathcal{E}(\pi)\right]-a\frac{\kappa}{2}d^2(\pi,\rho)+b+\frac{1}{2}a^2d^2(\pi,\rho)+abd(\pi,\rho)+\frac{1}{2}b^2.$$

Then the operator $H_\dagger\subseteq C(E)\times USC(E)$ is defined by

$$H_\dagger:=\left\{\left(f^\dagger_{a,b,c,\mu,\rho},g^\dagger_{a,b,c,\mu,\rho}\right)\,\middle|\,a,b>0,c\in\mathbb{R},\mu,\rho\in E:\,\mathcal{E}(\rho)<\infty\right\}.$$

[2] In this computation we use the formal bound $|\partial(f+g)|\geq||\partial f|-|\partial g||$. The local slope does not satisfy this property. In order to justify heuristically the upcoming calculations, it is convenient to think of it as a proxy for the norm of the gradient of $f+g$.

2. For each $a > 0, b > 0, c \in \mathbb{R}$, and $\pi, \gamma \in E : \mathcal{E}(\gamma) < \infty$ let $f^\ddagger = f^\ddagger_{a,b,c,\pi,\gamma} \in C(E)$ and $g^\ddagger = g^\ddagger_{a,b,c,\pi,\gamma} \in LSC(E)$ be given for any $\mu \in E$ by

$$f^\ddagger(\mu) := -\frac{1}{2} a d^2(\gamma, \mu) - b d_T(\mu, \pi) + c$$
$$g^\ddagger(\mu) := a\left[\mathcal{E}(\mu) - \mathcal{E}(\gamma)\right] + a\frac{\kappa}{2} d^2(\gamma, \mu) - b + \frac{1}{2} a^2 d^2(\gamma, \mu) - abd(\gamma, \mu) - \frac{1}{2} b^2.$$

Then the operator $H_\ddagger \subseteq C(E) \times LSC(E)$ is defined by

$$H_\ddagger := \left\{ \left(f^\ddagger_{a,b,c,\pi,\gamma}, g^\ddagger_{a,b,c,\pi,\gamma} \right) \,\middle|\, a, b > 0, c \in \mathbb{R}, \pi, \gamma \in E : \mathcal{E}(\gamma) < \infty \right\}.$$

Remark 2.12. Note that the term $-\frac{1}{2}b^2$, in the definition of $g^\ddagger$, is introduced in order to have more flexibility for an existence proof.

We are now ready to provide the notion of solution we are going to work with, which we state for general Hamiltonians $A_\dagger \subseteq LSC(E) \times USC(E)$ and $A_\ddagger \subseteq USC(E) \times LSC(E)$.

Definition 2.13. Fix $\lambda > 0$ and $h^\dagger, h^\ddagger \in C_b(E)$. Consider the equations

$$f - \lambda A_\dagger f = h^\dagger, \tag{2.10}$$
$$f - \lambda A_\ddagger f = h^\ddagger. \tag{2.11}$$

We say that u is a *(viscosity) subsolution* of equation (2.10) if u is bounded, upper semi-continuous and if for all $(f, g) \in A_\dagger$ there exists a sequence $(\pi_n)_{n\in\mathbb{N}} \in E$ such that

$$\lim_{n\uparrow\infty} u(\pi_n) - f(\pi_n) = \sup_\pi u(\pi) - f(\pi), \tag{2.12}$$
$$\limsup_{n\uparrow\infty} u(\pi_n) - \lambda g(\pi_n) - h^\dagger(\pi_n) \leq 0. \tag{2.13}$$

We say that v is a *(viscosity) supersolution* of equation (2.11) if v is bounded, lower semi-continuous and if for all $(f, g) \in A_\ddagger$ there exists a sequence $(\pi_n)_{n\in\mathbb{N}} \in E$ such that

$$\lim_{n\uparrow\infty} v(\pi_n) - f(\pi_n) = \inf_\pi v(\pi) - f(\pi),$$
$$\liminf_{n\uparrow\infty} v(\pi_n) - \lambda g(\pi_n) - h^\ddagger(\pi_n) \geq 0.$$

If $h^\dagger = h^\ddagger$, we say that u is a *(viscosity) solution* of equations (2.10) and (2.11) if it is both a subsolution of (2.10) and a supersolution of (2.11). We say that (2.10) and (2.11) satisfy the *comparison principle* if for every subsolution u to (2.10) and supersolution v to (2.11), we have $\sup_E u - v \leq \sup_E h^\dagger - h^\ddagger$.

In classical works on viscosity solutions, instead of working with the statement "there exists a sequence such that...", one has "for all optimizers one has...". However, when constructing our test functions in the comparison principle proof, we will work with the Ekeland variational principle, see Lemma A.1. This principle will give us an optimizer that is also unique. We will show in Lemma A.4 that, for our specific test functions, we can work directly with the unique optimizer instead of passing through an optimizing sequence as if we were using the stronger definition. On the other hand, Definition 2.13 is easier to handle while showing existence of solutions. We are ready to state the main result of this article.

Theorem 2.14. ***[The comparison Principle.]*** *Let Assumptions 2.3, 2.5 and 2.9 be satisfied. Let $\lambda > 0$ and $h^\dagger, h^\ddagger : E \to \mathbb{R}$ be bounded and uniformly continuous. Let $u : E \to \mathbb{R}$ be a viscosity subsolution to $f - \lambda H_\dagger f = h^\dagger$ and let $v : E \to \mathbb{R}$ be a viscosity supersolution to $f - \lambda H_\ddagger f = h^\ddagger$. Then we have*

$$\sup_{\pi \in E} u(\pi) - v(\pi) \leq \sup_{\pi \in E} h^\dagger(\pi) - h^\ddagger(\pi).$$

Remark 2.15. Note that we formally have

$$Hf \leq H_\dagger f \quad \text{and} \quad H_\ddagger f \leq Hf.$$

Thanks to these inequalities the above result will give a formal comparison principle for equation (2.1).

Remark 2.16. The assumption that $h^\dagger, h^\ddagger$ are uniformly continuous can be weakened to uniform continuity on sets of the type

$$K^\rho_{c,d} := \{\pi \in E \,|\, d(\pi, \rho) \leq c, \mathcal{E}(\rho) \leq d\}\,.$$

3. Proof of Theorem 2.14

The proof of Theorem 2.14 contains two main parts. The first part consists in showing that, in order to establish the comparison principle, we can reduce to the usual estimation on the difference of $H_\dagger$ and $H_\ddagger$. The estimation of this difference, however, is non-trivial in the present context and we postpone to section 3.2 the proof of some of the key estimates needed there.

Remark 3.1. In Step 1 of the proof below, we first make use of the fact that $\mathcal{E}$ can be bounded from below by a non-negative constant times $-d^2$. In this way, the standard quadruplication of variables, which goes with a penalization needed as we work with non-equal variables, is indeed a penalization. If $\mathcal{E}$ is itself already bounded from below by 0, we can simplify significantly the proof by choosing $c_1 = 0$.

Proof. Let u be a subsolution of equation (2.10) and v a supersolution of equation (2.11), we have to prove that

$$\sup_{\pi \in E} u(\pi) - v(\pi)$$

can be controlled by

$$\sup_{\pi \in E} h^{\dagger}(\pi) - h^{\ddagger}(\pi).$$

To proceed, as in the classical proof of the comparison theorem, one usually performs the doubling variables method, that can be done in our case using the distance function and the energy functional as penalization functions. However, the use of the energy functional and the fact that $\mathcal{E}(\pi)$ could be worth $+\infty$ oblige us to introduce two additional variables, i.e. we quadruplicate the number of variables. This procedure is actually reminiscent of the sup-convolution procedure.

Step 1: Quadruplication of variables and Ekeland's principle.
We fix $\nu_0 \in E$ such that $\mathcal{E}(\nu_0) < \infty$, we need $\mathcal{E}(\nu_0) < \infty$ and $c_1, c_2 \in \mathbb{R}$ as in Lemma 4.1 item (a), i.e. such that

$$\inf_{\pi \in E} \mathcal{E}(\pi) + \frac{c_1}{2} d^2(\pi, \nu_0) + c_2 = 0,$$

and we define

$$\bar{\mathcal{E}}(\pi) := \mathcal{E}(\pi) + \frac{c_1}{2} d^2(\pi, \nu_0) + c_2.$$

We fix $\alpha > 0$ and ε_α small enough (this value has to be fixed according to the condition (3.15), i.e. $\Xi_\alpha(x_{\alpha,0}) + \varepsilon_\alpha < \alpha^{-1}$, where $x_{\alpha,0} = (\pi_{\alpha,0}, \rho_{\alpha,0}, \mu_{\alpha,0}, \gamma_{\alpha,0})$ will be chosen later on and Ξ_α is defined as below). We introduce for $x = (\pi, \rho, \mu, \gamma) \in E^4$

$$\Phi_\alpha(x) := \frac{u(\pi)}{1 - \varepsilon_\alpha} - \frac{v(\mu)}{1 + \varepsilon_\alpha}$$

$$\Psi_\alpha(x) := \frac{d^2(\pi, \rho)}{2(1 - \varepsilon_\alpha)} + \frac{d^2(\rho, \gamma)}{2} + \frac{d^2(\gamma, \mu)}{2(1 + \varepsilon_\alpha)}$$

$$\Psi_{\alpha,0}(x) := \frac{1}{2(1 - \varepsilon_\alpha)} d^2(\pi, \mu)$$

$$\Xi_\alpha(x) := \frac{\varepsilon_\alpha}{1 - \varepsilon_\alpha} \bar{\mathcal{E}}(\rho) + \frac{\varepsilon_\alpha}{1 + \varepsilon_\alpha} \bar{\mathcal{E}}(\gamma)$$

Next, we define

$$\mathcal{G}_\alpha(x) := \Phi_\alpha(x) - \alpha \Psi_\alpha(x) - \Xi_\alpha(x), \quad M_\alpha := \sup_{x \in E^4} \mathcal{G}_\alpha(x) \tag{3.2a}$$

$$\mathcal{G}_{\alpha,0}(x) := \Phi_\alpha(x) - \alpha \Psi_{\alpha,0}(x), \quad M_{\alpha,0} := \sup_{x \in E^4} \mathcal{G}_{\alpha,0}(x)$$

and

$$\mathcal{B}(x,\tilde{x}) := \frac{1}{1-\varepsilon_\alpha} d_T(\pi,\tilde{\pi}) + \frac{1}{1+\varepsilon_\alpha} d_T(\mu,\tilde{\mu}) + d_T(\rho,\tilde{\rho}) + d_T(\gamma,\tilde{\gamma}).$$

We gather the important results of this step in the following proposition, whose proof is postponed to section 3.1.

Proposition 3.2. *For each $\alpha > 0$ we can find $x_\alpha = (\pi_\alpha, \rho_\alpha, \mu_\alpha, \gamma_\alpha) \in E^4$ such that*

(a)

$$\sup_{\pi \in E} u(\pi) - v(\pi) \leq \Phi_\alpha(x_\alpha) + \mathcal{O}(\alpha^{-1/2}), \tag{3.3}$$

(b) $\rho_\alpha, \gamma_\alpha \in \mathcal{D}(\mathcal{E})$ and x_α is the unique point in E^4 such that

$$\sup_{x \in E^4} \mathcal{G}_\alpha(x) - \frac{1}{2}\alpha^{-2} \leq \mathcal{G}_\alpha(x_\alpha) = \sup_{x \in E^4} \mathcal{G}_\alpha(x) - \alpha^{-1}\mathcal{B}_\alpha, \tag{3.4}$$

where

$$\mathcal{B}_\alpha(x) := \mathcal{B}(x, x_\alpha). \tag{3.5}$$

(c) If $(x_n)_{n\in\mathbb{N}} \in E^4$ is such that

$$\lim_{n\to\infty} \mathcal{G}_\alpha(x_n) - \alpha^{-1}\mathcal{B}_\alpha(x_n) = \mathcal{G}_\alpha(x_\alpha),$$

then $\lim_{n\to\infty} x_n = x_\alpha$.

(d) We have

$$\liminf_{\alpha\to\infty} \alpha\Psi_\alpha(x_\alpha) + \Xi_\alpha(x_\alpha) + \varepsilon_\alpha d^2(\rho_\alpha, \nu_0) + \varepsilon_\alpha d^2(\gamma_\alpha, \nu_0) = 0.$$

Step 2: Use of sub(super)solution properties. In the rest of the proof we consider a diverging sequence $(\alpha_n)_{n\in\mathbb{N}}$ along which

$$\lim_{n\to\infty} \alpha_n \Psi_{\alpha_n}(x_{\alpha_n}) + \Xi_{\alpha_n}(x_{\alpha_n}) + \varepsilon_{\alpha_n} d^2(\rho_{\alpha_n}, \nu_0) + \varepsilon_{\alpha_n} d^2(\gamma_{\alpha_n}, \nu_0) = 0.$$

Consider as test functions $f^\dagger, f^\ddagger : E \to (-\infty, +\infty)$ given by

$$f^\dagger(\cdot) := -(1-\varepsilon_{\alpha_n})\mathcal{G}_{\alpha_n}(\cdot, \mu_{\alpha_n}, \rho_{\alpha_n}, \gamma_{\alpha_n}) + u(\cdot) + (1-\varepsilon_{\alpha_n})\alpha_n^{-1}\mathcal{B}_{\alpha_n}(\cdot, \mu_{\alpha_n}, \rho_{\alpha_n}, \gamma_{\alpha_n}), \tag{3.6}$$

$$f^\ddagger(\cdot) := (1+\varepsilon_{\alpha_n})\mathcal{G}_{\alpha_n}(\pi_{\alpha_n}, \cdot, \rho_{\alpha_n}, \gamma_{\alpha_n}) + v(\cdot) - (1+\varepsilon_{\alpha_n})\alpha_n^{-1}\mathcal{B}_{\alpha_n}(\pi_{\alpha_n}, \cdot, \rho_{\alpha_n}, \gamma_{\alpha_n}).$$

Note that $f^\dagger, f^\ddagger$ are valid test functions. Indeed, from (3.2a), (3.5) we have

$$f^\dagger(\pi) = \frac{\alpha_n}{2} d^2(\pi, \rho_{\alpha_n}) + \alpha_n^{-1} d_T(\pi, \pi_{\alpha_n}) + \text{const.},$$
$$f^\ddagger(\mu) = -\frac{\alpha_n}{2} d^2(\mu, \gamma_{\alpha_n}) - {\alpha_n}^{-1} d_T(\mu, \mu_{\alpha_n}) + \text{const.},$$

and we know that $\rho_{\alpha_n}, \gamma_{\alpha_n} \in \mathcal{D}(\mathcal{E})$ by Proposition 3.2-(b). From the very definition of $f^\dagger$, we obtain

$$u(\pi) - f^\dagger(\pi) = (1 - \varepsilon_{\alpha_n})[\mathcal{G}_{\alpha_n} - {\alpha_n}^{-1}\mathcal{B}_{\alpha_n}](\pi, \mu_{\alpha_n}, \rho_{\alpha_n}, \gamma_{\alpha_n}), \tag{3.7}$$

and π_{α_n} is the unique maximizer of $u(\pi) - f^\dagger(\pi)$ because of (3.4). Analogously, we find

$$v(\mu) - f^\ddagger(\mu) = -(1 + \varepsilon_{\alpha_n})[\mathcal{G}_{\alpha_n} - {\alpha_n}^{-1}\mathcal{B}_{\alpha_n}](\pi_{\alpha_n}, \mu, \rho_{\alpha_n}, \gamma_{\alpha_n}),$$

and μ_{α_n} is the unique minimizer of $v(\mu) - f^\ddagger(\mu)$. Being u a subsolution, there exists a sequence $(\pi_m)_{m\in\mathbb{N}} \in E$ satisfying (2.12) and (2.13), for $(f^\dagger, g^\dagger) \in H_\dagger$, where $g^\dagger$ is given by Definition 2.11 (with $a = \alpha_n, b = {\alpha_n}^{-1}$). In the next lines, we deduce from these properties that

$$u(\pi_{\alpha_n}) \le \lambda g^\dagger(\pi_{\alpha_n}) + h^\dagger(\pi_{\alpha_n}). \tag{3.8}$$

We begin by observing that

$$\begin{aligned}
\lim_{m\to+\infty} (1 - \varepsilon_{\alpha_n})[\mathcal{G}_{\alpha_n} - {\alpha_n}^{-1}\mathcal{B}_{\alpha_n}](\pi_m, \mu_{\alpha_n}, \rho_{\alpha_n}, \gamma_{\alpha_n}) &\overset{(3.7)}{=} \lim_{m\to\infty} \left(u - f^\dagger\right)(\pi_m) \\
&\overset{(2.12)}{=} \sup_{\pi\in E} \left(u - f^\dagger\right)(\pi) \\
&\overset{(3.7)+(3.4)}{=} (1 - \varepsilon_{\alpha_n})\mathcal{G}_{\alpha_n}(x_{\alpha_n}) \\
&= (u - f^\dagger)(\pi_{\alpha_n}).
\end{aligned}$$

At this point, we can use item (c) of Proposition 3.2 which gives that $\lim_{m\to+\infty} \pi_m = \pi_{\alpha_n}$. Now Lemma A.4, says that since there exists $\pi_{\alpha_n} \in E$ such that $\lim_{m\to+\infty} \pi_m = \pi_{\alpha_n}$ and

$$u(\pi_{\alpha_n}) - f^\dagger(\pi_{\alpha_n}) = \sup_\pi u(\pi) - f^\dagger(\pi).$$

Then we have

$$u(\pi_{\alpha_n}) - \lambda g^\dagger(\pi_{\alpha_n}) - h^\dagger(\pi_{\alpha_n}) \le 0.$$

Therefore we finally establish (3.8). Arguing similarly, we obtain that

$$v(\mu_{\alpha_n}) \ge \lambda g^\ddagger(\mu_{\alpha_n}) + h^\ddagger(\mu_{\alpha_n}),$$

for $(f^{\ddagger}, g^{\ddagger}) \in H_{\ddagger}$, where $g^{\ddagger}$ is given by Definition 2.11 (with $a = \alpha_n, b = {\alpha_n}^{-1}$). Plugging (3.8) and this last bound into (3.3) and using the fact that our choice (3.14) of ε_{α_n} implies $\varepsilon_{\alpha_n} \leq \alpha_n^{-1}$, we arrive at

$$\begin{aligned}\sup_{\pi\in E} u(\pi) - v(\pi) \leq& \frac{h^{\dagger}(\pi_{\alpha_n})}{1-\varepsilon_{\alpha_n}} - \frac{h^{\ddagger}(\mu_{\alpha_n})}{1+\varepsilon_{\alpha_n}} + \lambda\Big(\frac{1}{1-\varepsilon_{\alpha_n}} g^{\dagger}(\pi_{\alpha_n}) - \frac{1}{1+\varepsilon_{\alpha_n}} g^{\ddagger}(\mu_{\alpha_n})\Big) \\ &+ \mathcal{O}({\alpha_n}^{-1/2}).\end{aligned} \tag{3.9}$$

Step 3: Upper bound on the difference of the Hamiltonians. Applying the definition of $g^{\dagger}$ and $g^{\ddagger}$ and with the help of Proposition 3.2 (d) we can split the difference of the Hamiltoniains into two terms and a vanishing term, namely

$$\begin{aligned}&\frac{g^{\dagger}(\pi_{\alpha_n})}{1-\varepsilon_{\alpha_n}} - \frac{g^{\ddagger}(\mu_{\alpha_n})}{1+\varepsilon_{\alpha_n}} \leq \\ &\alpha_n\Big[\frac{1}{1-\varepsilon_{\alpha_n}}\big(\mathcal{E}(\rho_{\alpha_n}) - \mathcal{E}(\pi_{\alpha_n}) + \frac{\kappa}{2} d^2(\pi_{\alpha_n}, \rho_{\alpha_n})\big) \\ &- \frac{1}{1+\varepsilon_{\alpha_n}}\big(\mathcal{E}(\mu_{\alpha_n}) - \mathcal{E}(\gamma_{\alpha_n}) - \frac{\kappa}{2} d^2(\pi_{\alpha_n}, \rho_{\alpha_n})\big)\Big] \\ &+ \frac{{\alpha_n}^2}{2(1-\varepsilon_{\alpha_n})} d^2(\pi_{\alpha_n}, \rho_{\alpha_n}) - \frac{{\alpha_n}^2}{2(1+\varepsilon_{\alpha_n})} d^2(\mu_{\alpha_n}, \gamma_{\alpha_n}) \\ &+ o(1)\end{aligned} \tag{3.10}$$

We gather here the important estimates used in this step and that will be contained in Lemma 3.3 and Lemma 3.4, whose proof is postponed to section 3.2.

Let $(\alpha_n)_{n\in\mathbb{N}}$ be the sequence given by Proposition 3.2-(d), then we have

$$\begin{aligned}&\alpha_n\Big[\frac{1}{1-\varepsilon_{\alpha_n}}\big(\mathcal{E}(\rho_{\alpha_n}) - \mathcal{E}(\pi_{\alpha_n}) + \frac{\kappa}{2} d^2(\pi_{\alpha_n}, \rho_{\alpha_n})\big) \\ &- \frac{1}{1+\varepsilon_{\alpha_n}}\big(\mathcal{E}(\mu_{\alpha_n}) - \mathcal{E}(\gamma_{\alpha_n}) - \frac{\kappa}{2} d^2(\pi_{\alpha_n}, \rho_{\alpha_n})\big)\Big] \\ &\leq -\frac{\varepsilon_{\alpha_n}}{(1-\varepsilon_{\alpha_n})} I(\rho_{\alpha_n}) - \frac{\varepsilon_{\alpha_n}}{(1+\varepsilon_{\alpha_n})} I(\gamma_{\alpha_n}) + o(1),\end{aligned} \tag{3.11}$$

and

$$\begin{aligned}&\frac{\alpha_n^2}{2(1-\varepsilon_{\alpha_n})} d^2(\pi_{\alpha_n}, \rho_{\alpha_n}) - \frac{\alpha_n^2}{2(1+\varepsilon_{\alpha_n})} d^2(\mu_{\alpha_n}, \gamma_{\alpha_n}) \\ &\leq \frac{\varepsilon_{\alpha_n}}{(1-\varepsilon_{\alpha_n})} I(\rho_{\alpha_n}) + \frac{\varepsilon_{\alpha_n}}{(1+\varepsilon_{\alpha_n})} I(\gamma_{\alpha_n}) + o(1).\end{aligned} \tag{3.12}$$

If we now apply (3.11) to bound the first term and (3.12) to bound the second term, we obtain that

$$\frac{g^\dagger(\pi_{\alpha_n})}{1-\varepsilon_{\alpha_n}} - \frac{g^\ddagger(\mu_{\alpha_n})}{1+\varepsilon_{\alpha_n}} \leq o(1).$$

Step 4: Conclusion. Let $\omega^\dagger$ be a modulus of continuity for $h^\dagger$. Combining the conclusion of *Step 3* with (3.9) we obtain that for all $n \in \mathbb{N}$

$$\begin{aligned}\sup_{\pi\in E} u(\pi) - v(\pi) &\leq \omega^\dagger(d(\pi_{\alpha_n}, \mu_{\alpha_n})) + \frac{h^\dagger(\mu_{\alpha_n})}{1-\varepsilon_{\alpha_n}} - \frac{h^\ddagger(\mu_{\alpha_n})}{1+\varepsilon_{\alpha_n}} + o(1) \\ &\leq \sup_{\pi\in E} h^\dagger(\pi) - h^\ddagger(\pi) + \omega^\dagger(d(\pi_{\alpha_n}, \mu_{\alpha_n})) + o(1)\end{aligned}$$

where to establish the last inequality we used the boundedness of $h^\dagger, h^\ddagger$ and (3.14). The desired conclusion follows by taking limits on both sides in the above display and invoking one last time Proposition 3.2(d) Note that item (d) of Proposition 3.2 also implies Remark 2.16. □

3.1. Proof of Proposition 3.2

Proof.
- *Step 1: quadruplication of variables* We first pick $(\pi_{\alpha,0}, \mu_{\alpha,0}) \in E^2$ such that

$$\sup_{\pi\in E} u(\pi) - v(\pi) \leq u(\pi_{\alpha,0}) - v(\mu_{\alpha,0}) - \frac{\alpha}{2} d^2(\pi_{\alpha,0}, \mu_{\alpha,0}) + \alpha^{-1}. \tag{3.13}$$

  Next, we choose $(\rho_{\alpha,0}, \gamma_{\alpha,0}) \in E^2$ such that

$$\mathcal{E}(\rho_{\alpha,0}) + \mathcal{E}(\gamma_{\alpha,0}) < +\infty, \quad d(\pi_{\alpha,0}, \rho_{\alpha,0}) + d(\gamma_{\alpha,0}, \mu_{\alpha,0}) < \alpha^{-1}, \tag{3.14}$$

  and $\varepsilon_\alpha \in (0, 1/3)$ such that

$$\Xi_\alpha(x_{\alpha,0}) + \varepsilon_\alpha < \alpha^{-1}, \tag{3.15}$$

  where $x_{\alpha,0} = (\pi_{\alpha,0}, \rho_{\alpha,0}, \mu_{\alpha,0}, \gamma_{\alpha,0})$.
- *Step 2: algebraic bounds on the difference of solutions* In this step we show that

$$\sup_{\pi\in E} u(\pi) - v(\pi) \leq M_{\alpha,0} + \mathcal{O}(\alpha^{-1}) \leq M_\alpha + \mathcal{O}(\alpha^{-1/2}). \tag{3.16}$$

  We do so by first showing that

$$\begin{aligned}\sup_{\pi\in E} u(\pi) - v(\pi) &\leq \sup_{x\in E^4} \Phi_\alpha(x) - \alpha\Psi_{\alpha,0}(x) + \mathcal{O}(\alpha^{-1}) \\ &\leq \Phi_\alpha(x_{\alpha,0}) - \alpha\Psi_{\alpha,0}(x_{\alpha,0}) + \mathcal{O}(\alpha^{-1})\end{aligned} \tag{3.17}$$

  and eventually establishing that

$$\alpha\Psi_{\alpha,0}(x_{\alpha,0}) \geq \alpha\Psi_\alpha(x_{\alpha,0}) + \mathcal{O}(\alpha^{-1/2}). \tag{3.18}$$

Once these two bounds have been proven, the desired conclusion (3.16) follows at once using (3.15). Let us now proceed to the proof of (3.17). From the boundedness of u, v and using the bounds

$$\left|\frac{1}{1-\varepsilon_\alpha} - 1\right| + \left|\frac{1}{1+\varepsilon_\alpha} - 1\right| = \mathcal{O}(\varepsilon_\alpha) = \mathcal{O}(\alpha^{-1}) \tag{3.19}$$

we get

$$\sup_{\pi\in E} u(\pi) - v(\pi) \leq \Phi_\alpha(x_{\alpha,0}) - \frac{\alpha}{2} d^2(\pi_{\alpha,0}, \mu_{\alpha,0}) + \mathcal{O}(\alpha^{-1}) \tag{3.20}$$

From the choice of $(\pi_{\alpha,0}, \mu_{\alpha,0})$ (see (3.13)) we deduce that

$$\frac{\alpha}{2} d^2(\pi_{\alpha,0}, \mu_{\alpha,0}) \leq \sup_\pi |u|(\pi) + \sup_\pi |v|(\pi) + \sup_\pi |u - v|(\pi) + \alpha^{-1} = \mathcal{O}(1) \tag{3.21}$$

But then, using this last bound and (3.19) in (3.20) we obtain

$$\sup_{\pi\in E} u(\pi) - v(\pi) \leq \Phi_\alpha(x_{\alpha,0}) - \alpha\Psi_{\alpha,0}(x_{\alpha,0}) + \mathcal{O}(\alpha^{-1}),$$

which proves the first inequality of (3.17). To prove the second one, i.e.

$$\sup_{x\in E^4} \Phi_\alpha(x) - \alpha\Psi_{\alpha,0}(x) + \mathcal{O}(\alpha^{-1}) \leq \Phi_\alpha(x_{\alpha,0}) - \alpha\Psi_{\alpha,0}(x_{\alpha,0}) + \mathcal{O}(\alpha^{-1}),$$

we proceed as before using the boundedness of u, v, (3.19) and (3.21), to show that

$$\sup_{x\in E^4} \Phi_\alpha(x) - \alpha\Psi_{\alpha,0}(x) + \mathcal{O}(\alpha^{-1}) \leq \sup_{\pi,\mu\in E} u(\pi) - v(\mu) - \frac{\alpha}{2} d^2(\pi, \mu) + \mathcal{O}(\alpha^{-1}).$$

By the choice of $(\pi_{\alpha,0}, \mu_{\alpha,0})$ (see (3.13)) we obtain

$$\sup_{x\in E^4} \Phi_\alpha(x) - \alpha\Psi_{\alpha,0}(x) + \mathcal{O}(\alpha^{-1}) \leq u(\pi_{\alpha,0}) - v(\mu_{\alpha,0}) - \frac{\alpha}{2} d^2(\pi_{\alpha,0}, \mu_{\alpha,0}) + \mathcal{O}(\alpha^{-1}),$$

and, through analogous computations, the second inequality of (3.17). In order to prove (3.18) we begin observing that the triangular inequality give

$$d(\pi_{\alpha,0}, \mu_{\alpha,0}) \geq d(\rho_{\alpha,0}, \gamma_{\alpha,0}) - d(\pi_{\alpha,0}, \rho_{\alpha,0}) - d(\gamma_{\alpha,0}, \mu_{\alpha,0}). \tag{3.22}$$

There are two possible cases:

– $d(\rho_{\alpha,0}, \gamma_{\alpha,0}) < d(\pi_{\alpha,0}, \rho_{\alpha,0}) + d(\gamma_{\alpha,0}, \mu_{\alpha,0})$. In this case, we immediately obtain from our choice of $\rho_{\alpha,0}$ and $\gamma_{\alpha,0}$ that $d(\rho_{\alpha,0}, \gamma_{\alpha,0}) = \mathcal{O}(\alpha^{-1})$ from which we deduce that

$$d^2(\pi_{\alpha,0}, \mu_{\alpha,0}) = d^2(\pi_{\alpha,0}, \rho_{\alpha,0}) + d^2(\rho_{\alpha,0}, \gamma_{\alpha,0}) + d^2(\gamma_{\alpha,0}, \mu_{\alpha,0}) + \mathcal{O}(\alpha^{-2}).$$

– $d(\rho_{\alpha,0},\gamma_{\alpha,0}) \geq d(\pi_{\alpha,0},\rho_{\alpha,0}) + d(\gamma_{\alpha,0},\mu_{\alpha,0})$. In this case taking squares in (3.22) and using (3.14) we get

$$\begin{aligned} d^2(\pi_{\alpha,0},\mu_{\alpha,0}) =& d^2(\pi_{\alpha,0},\rho_{\alpha,0}) + d^2(\rho_{\alpha,0},\gamma_{\alpha,0}) + d^2(\gamma_{\alpha,0},\mu_{\alpha,0}) \\ &+ d(\rho_{\alpha,0},\gamma_{\alpha,0})\mathcal{O}(\alpha^{-1}) + \mathcal{O}(\alpha^{-2}). \end{aligned}$$

An application of the triangular inequality (3.22) in combination with (3.14) and (3.21) gives that $d(\rho_{\alpha,0},\gamma_{\alpha,0}) = \mathcal{O}(\alpha^{-1/2})$. Plugging this into the above display yields

$$d^2(\pi_{\alpha,0},\mu_{\alpha,0}) = d^2(\pi_{\alpha,0},\rho_{\alpha,0}) + d^2(\rho_{\alpha,0},\gamma_{\alpha,0}) + d^2(\gamma_{\alpha,0},\mu_{\alpha,0}) + \mathcal{O}(\alpha^{-3/2}). \quad (3.23)$$

Therefore in both cases we have that (3.23) holds. Multiplying this relation on both sides by $\frac{\alpha}{2(1-\varepsilon_\alpha)}$ and using the basic inequality $\frac{\alpha}{2(1-\varepsilon_\alpha)} \leq \frac{\alpha}{2} \leq \frac{\alpha}{2(1+\varepsilon_\alpha)}$ establishes (3.18).

- *Step 3: Ekeland's principle and proof of item (a),(b) and (c)* The verification that $\mathcal{G}_\alpha$ and $\mathcal{B}$ satisfy the hypothesis of Ekeland's Lemma (Lemma A.1) is done separately in Lemma A.3 in the Appendix. Next, we pick $\hat{x} = (\hat{\pi},\hat{\mu},\hat{\rho},\hat{\gamma}) \in E^2 \times \mathcal{D}(\mathcal{E})^2$ such that

$$\sup_{x\in E^4} \mathcal{G}_\alpha(x) - \frac{1}{2}\alpha^{-2} \leq \mathcal{G}_\alpha(\hat{x}). \quad (3.24)$$

If we now apply Lemma A.1 setting $\delta = \alpha^{-1}$ we immediately obtain the equality statement in (3.4) thanks to item (2)-A.1. I.e., for each $\alpha > 0$ we can find a unique $x_\alpha = (\pi_\alpha,\rho_\alpha,\mu_\alpha,\gamma_\alpha) \in E^2 \times \mathcal{D}(\mathcal{E}^2)$ that attains the maximum in $\sup_{E^4} \mathcal{G}_\alpha(\cdot) - \alpha^{-1}\mathcal{B}_\alpha(\cdot)$. Moreover, using item (1)-A.1 in combination with (3.24) we prove the inequality statement in (3.4). This concludes the proof of item (b). At this point, item (a) is a direct consequence of equations (3.4), that we have just proven, (3.16), and the fact that Ξ_α,Ψ_α non-negative functions. Item (c) also follows from item (c)-A.1.
- *Step 4: Proof of item (d).* We have from item (b)

$$M_\alpha - \frac{1}{2}\alpha^{-2} \leq \mathcal{G}_\alpha(x_\alpha) \overset{\Xi_\alpha\geq 0}{\leq} [\Phi_\alpha - \alpha\Psi_\alpha](x_\alpha). \quad (3.25)$$

Next, we observe that our choice of ε_α and the boundedness of u,v imply

$$\Phi_\alpha(x_\alpha) = \Phi_{\alpha/6}(x_\alpha) + \mathcal{O}(\alpha^{-1}).$$

Moreover, using the version of Jensen's inequality (A.3), proven separately in Lemma A.5, with the choices $\varepsilon = \varepsilon_\alpha, \varepsilon' = \varepsilon_{\alpha/6}$ we obtain

$$\Psi_\alpha(x_\alpha) \geq \frac{1}{3}\Psi_{\frac{\alpha}{6},0}(x_\alpha).$$

But then, the right hand side in (3.25) is bounded above by

$$\begin{aligned} M_\alpha - \frac{1}{2}\alpha^{-2} &\leq [\Phi_{\alpha/6} - \frac{\alpha}{6}\Psi_{\alpha/6,0}](x_\alpha) - \frac{\alpha}{2}\Psi_\alpha(x_\alpha) + \mathcal{O}(\alpha^{-1}) \\ &\leq M_{\alpha/6,0} - \frac{\alpha}{2}\Psi_\alpha(x_\alpha) + \mathcal{O}(\alpha^{-1}) \\ &\overset{(3.16)}{\leq} M_{\alpha/6} - \frac{\alpha}{2}\Psi_\alpha(x_\alpha) + \mathcal{O}(\alpha^{-1/2}). \end{aligned}$$

We have thus obtained

$$M_\alpha - \alpha^{-2} \leq M_{\alpha/6} - \frac{\alpha}{2}\Psi_\alpha(x_\alpha) + \mathcal{O}(\alpha^{-1/2}).$$

Taking lim sup on both sides we get

$$\limsup_{\alpha\to\infty} M_\alpha \leq \limsup_{\alpha\to\infty} M_{\alpha/6} - \frac{\alpha}{2}\Psi_\alpha(x_\alpha) \leq \limsup_{\alpha\to\infty} M_{\alpha/6} - \frac{1}{2}\liminf_{\alpha\to\infty} \alpha\Psi_\alpha(x_\alpha),$$

whence the existence of a sequence $(\alpha_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to+\infty} \alpha_n = +\infty, \quad \lim_{n\to+\infty} \alpha_n\Psi_{\alpha_n}(x_{\alpha_n}) = 0. \tag{3.26}$$

To conclude the proof, we observe that thanks to item (b) we have

$$\mathcal{G}_\alpha(x_\alpha) \geq \mathcal{G}_\alpha(x_{\alpha,0}) - \frac{1}{2}\alpha^{-2},$$

whence, with the help of (3.15)

$$\Xi_\alpha(x_\alpha) \leq [\Phi_\alpha - \alpha\Psi_\alpha](x_\alpha) - [\Phi_\alpha - \alpha\Psi_\alpha](x_{\alpha,0}) + \mathcal{O}(\alpha^{1/2}).$$

Using (3.18) on $\alpha\Psi_\alpha(x_{\alpha,0})$ and Lemma A.5 to obtain $-\alpha\Psi_\alpha(x_\alpha) \leq -\alpha\Psi_{\alpha,0}(x_\alpha) + 5\alpha\Psi_\alpha(x_\alpha)$, we obtain

$$\Xi_\alpha(x_\alpha) \leq \mathcal{G}_{\alpha,0}(x_\alpha) + 5\alpha\Psi_\alpha(x_\alpha) - \mathcal{G}_{\alpha,0}(x_{\alpha,0}) + \mathcal{O}(\alpha^{-1/2}).$$

Since $\mathcal{G}_{\alpha,0}(x_\alpha) \leq M_{\alpha,0}$ and $\mathcal{G}_{\alpha,0}(x_{\alpha,0}) = M_{\alpha,0} + \mathcal{O}(\alpha^{-1})$ by (3.17), we find

$$\Xi_\alpha(x_\alpha) \leq 5\alpha\Psi_\alpha(x_\alpha) + \mathcal{O}(\alpha^{-1/2}).$$

As a consequence of (3.26), if we choose the same sequence $(\alpha_n)_{n\in\mathbb{N}}$ giving (3.26) we have

$$\lim_{n\to+\infty} \Xi_{\alpha_n}(x_{\alpha_n}) = 0. \tag{3.27}$$

Finally, observing that by construction

$$\bar{\mathcal{E}}(\pi) := \mathcal{E}(\pi) + \frac{c_1}{2} d^2(\pi, \nu_0) + c_2,$$

for all $\pi \in E$ and some $c_1 > 0, c_2 \in \mathbb{R}$ this implies by Lemma 4.1 (a) that

$$\bar{\mathcal{E}}(\pi) \geq \tilde{c}_1 d^2(\pi, \nu_0) + \tilde{c}_2$$

for all $\pi \in E$ and some $\tilde{c}_1 > 0, \tilde{c}_2 \in \mathbb{R}$, we deduce from (3.27) that

$$\lim_{n \to +\infty} \varepsilon_{\alpha_n} d^2(\rho_{\alpha_n}, \nu_0) + \varepsilon_{\alpha_n} d^2(\gamma_{\alpha_n}, \nu_0) = 0. \quad \square$$

3.2. Key estimates

We now prove the two main estimates we used in the proof of the comparison principle. In the next lemma, we find an upper bound for the first term on the right-hand side in (3.10) relying essentially on (EVI_κ). It is precisely here where the use of d instead of d_T in Ekeland's lemma results in weaker estimates that do not allow to conclude the proof of the comparison principle. In Lemma 3.4, we find an upper bound for the second term on the right-hand side in (3.10), relying on the curves introduced in Assumption 2.9. The proofs of these lemmas are partially inspired by Lemma 2.5 and 2.6 of [30]. In both statements, we use the information functional $I = |\partial \mathcal{E}|^2$ which was introduced in Definition 2.8.

Lemma 3.3 *(Estimate on drift from EVI and gradient flow). For fixed $\alpha > 0$ let $x_\alpha = (\pi_\alpha, \mu_\alpha, \rho_\alpha, \gamma_\alpha)$ and ν_0 be as in the proof of Theorem 2.14. Then, we have that $\mathcal{E}(\pi_\alpha) + \mathcal{E}(\mu_\alpha) < +\infty$ and the following estimates hold*

$$\begin{aligned}
&\alpha\,[\mathcal{E}(\rho_\alpha) - \mathcal{E}(\pi_\alpha)] + \frac{\alpha\kappa}{2} d^2(\rho_\alpha, \pi_\alpha) \\
&\leq (1-\varepsilon_\alpha)\alpha\,[\mathcal{E}(\gamma_\alpha) - \mathcal{E}(\rho_\alpha)] - (1-\varepsilon_\alpha)\alpha\frac{\kappa}{2} d^2(\rho_\alpha, \gamma_\alpha) - \varepsilon_\alpha I(\rho_\alpha) + (1-\varepsilon_\alpha)\alpha^{-1} \\
&\quad + \varepsilon_\alpha c_1 [\mathcal{E}(\nu_0) - \mathcal{E}(\rho_\alpha)] - \varepsilon_\alpha c_1 \frac{\kappa}{2} d^2(\rho_\alpha, \nu_0);
\end{aligned} \tag{3.28}$$

$$\begin{aligned}
&\alpha\,[\mathcal{E}(\mu_\alpha) - \mathcal{E}(\gamma_\alpha)] - \frac{\alpha\kappa}{2} d^2(\gamma_\alpha, \mu_\alpha) \\
&\geq (1+\varepsilon_\alpha)\alpha\,[\mathcal{E}(\gamma_\alpha) - \mathcal{E}(\rho_\alpha)] + (1+\varepsilon_\alpha)\alpha\frac{\kappa}{2} d^2(\rho_\alpha, \gamma_\alpha) + \varepsilon_\alpha I(\gamma_\alpha) - (1+\varepsilon_\alpha)\alpha^{-1} \\
&\quad + \varepsilon_\alpha c_1 [\mathcal{E}(\gamma_\alpha) - \mathcal{E}(\nu_0)] + \varepsilon_\alpha c_1 \frac{\kappa}{2} d^2(\gamma_\alpha, \nu_0).
\end{aligned} \tag{3.29}$$

Moreover, $I(\rho_\alpha) + I(\gamma_\alpha) < \infty$. As a corollary, if $(\alpha_n)_{n \in \mathbb{N}}$ is the sequence given by Proposition 3.2-(d), then we have

$$\begin{aligned}&\alpha_n\Big[\frac{1}{1-\varepsilon_{\alpha_n}}\big(\mathcal{E}(\rho_{\alpha_n})-\mathcal{E}(\pi_{\alpha_n})+\frac{\kappa}{2}d^2(\pi_{\alpha_n},\rho_{\alpha_n})\big)\\&-\frac{1}{1+\varepsilon_{\alpha_n}}\big(\mathcal{E}(\mu_{\alpha_n})-\mathcal{E}(\gamma_{\alpha_n})-\frac{\kappa}{2}d^2(\pi_{\alpha_n},\rho_{\alpha_n})\big)\Big]\\&\leq -\frac{\varepsilon_{\alpha_n}}{(1-\varepsilon_{\alpha_n})}I(\rho_{\alpha_n})-\frac{\varepsilon_{\alpha_n}}{(1+\varepsilon_{\alpha_n})}I(\gamma_{\alpha_n})+o(1).\end{aligned} \tag{3.30}$$

Proof. The fact that $\mathcal{E}(\pi_\alpha)<+\infty$ follows from the subsolution property (3.8) of u and the fact that $u(\pi_\alpha),h^\dagger(\pi_\alpha),\mathcal{E}(\rho_\alpha)$ are all finite quantities. The proof that $\mathcal{E}(\mu_\alpha)<+\infty$ is analogous. Fix $s>0$. From (EVI_κ) and Ekeland's principle (3.4) we obtain that the gradient flow started at ρ_α satisfies

$$\begin{aligned}&\alpha\int_0^s \mathcal{E}(\rho_\alpha(r))-\mathcal{E}(\pi_\alpha)+\frac{\kappa}{2}d^2(\rho_\alpha(r),\pi_\alpha)\,dr\\&\leq \frac{\alpha d^2(\rho_\alpha,\pi_\alpha)}{2}-\frac{\alpha d^2(\rho_\alpha(s),\pi_\alpha)}{2}\\&=(1-\varepsilon_\alpha)\Big(\frac{\alpha d^2(\rho_\alpha,\pi_\alpha)}{2(1-\varepsilon_\alpha)}+\mathcal{G}_\alpha(x_\alpha)\Big)-(1-\varepsilon_\alpha)\Big(\frac{\alpha d^2(\rho_\alpha(s),\pi_\alpha)}{2(1-\varepsilon_\alpha)}+\mathcal{G}_\alpha(x_\alpha)\Big)\\&\leq (1-\varepsilon_\alpha)\Big(\frac{\alpha d^2(\rho_\alpha,\pi_\alpha)}{2(1-\varepsilon_\alpha)}+\mathcal{G}_\alpha(x_\alpha)\Big)\\&-(1-\varepsilon_\alpha)\Big(\frac{\alpha d^2(\rho_\alpha(s),\pi_\alpha)}{2(1-\varepsilon_\alpha)}+\mathcal{G}_\alpha(\pi_\alpha,\mu_\alpha,\rho_\alpha(s),\gamma_\alpha)-\alpha^{-1}\mathcal{B}_\alpha(\pi_\alpha,\mu_\alpha,\rho_\alpha(s),\gamma_\alpha)\Big).\end{aligned}$$

Recalling (3.2a), we can rewrite the last expression as

$$(1-\varepsilon_\alpha)\alpha\left[\frac{d^2(\rho_\alpha(s),\gamma_\alpha)}{2}-\frac{d^2(\rho_\alpha,\gamma_\alpha)}{2}\right] \tag{3.31}$$

$$+\varepsilon_\alpha[\mathcal{E}(\rho_\alpha(s))-\mathcal{E}(\rho_\alpha)]+\varepsilon_\alpha\frac{c_1}{2}[d^2(\rho_\alpha(s),\nu_0)-d^2(\rho_\alpha,\nu_0)] \tag{3.32}$$

$$+(1-\varepsilon_\alpha)\alpha^{-1}d_T(\rho_\alpha(s),\rho_\alpha). \tag{3.33}$$

Using (EVI_κ) in (3.31), the energy identity (4.2), again (EVI_κ) in (3.32) and Lemma 4.3 (b) in (3.33) we obtain the upper bound

$$\begin{aligned}&\int_0^s \alpha(1-\varepsilon_\alpha)[\mathcal{E}(\gamma_\alpha)-\mathcal{E}(\rho_\alpha(r))-\frac{\kappa}{2}d^2(\rho_\alpha(r),\gamma_\alpha)]-\varepsilon_\alpha I(\rho_\alpha(r))dr+(1-\varepsilon_\alpha)\alpha^{-1}s\\&+\int_0^s \varepsilon_\alpha c_1[\mathcal{E}(\nu_0)-\mathcal{E}(\rho_\alpha(r))-\frac{\kappa}{2}d^2(\rho_\alpha(r),\nu_0)]dr.\end{aligned}$$

Dividing by s and letting $s\to 0$ we obtain (3.28), recalling that $r\mapsto d^2(\rho_\alpha(r),\gamma_\alpha)$, $r\mapsto d^2(\rho_\alpha(r),\nu_0)$, $r\mapsto \mathcal{E}(\rho_\alpha(r))$ are continuous functions and that $r\mapsto I(\rho_\alpha(r))$ is right continuous by Lemma 4.1 (d). Arguing in the same way, we obtain (3.29). Finally, having proved (3.28), if we observe that all terms except $I(\rho_\alpha)$ are finite, we can deduce that $I(\rho_\alpha) < +\infty$. The proof that $I(\gamma_\alpha) < +\infty$ is completely analogous. At this point, inequality (3.30) follows due to Proposition 3.2-(d). □

In the following lemma we obtain an upper bound for the second term in (3.10). Here, it is the fact that (E, d) is a geodesic space together with the geometric conditions (2.6) (2.7) that play a crucial role.

Lemma 3.4. *For fixed $\alpha > 0$, let $x_\alpha = (\pi_\alpha, \mu_\alpha, \rho_\alpha, \gamma_\alpha)$ be as in the proof of Theorem 2.14. Then we have*

$$\frac{\alpha^2}{2} d^2(\rho_\alpha, \pi_\alpha) \leq (1 - \varepsilon_\alpha)\frac{1}{2}[\alpha^{-1} + \alpha d(\rho_\alpha, \gamma_\alpha)]^2 + \frac{\varepsilon_\alpha}{2}[\sqrt{I(\rho_\alpha)} + c_1 d(\rho_\alpha, \nu_0)]^2 \tag{3.34}$$

and

$$\frac{\alpha^2}{2} d^2(\gamma_\alpha, \mu_\alpha) \geq (1 + \varepsilon_\alpha)\frac{\alpha^2}{2} d^2(\gamma_\alpha, \rho_\alpha) - \frac{1}{2}\varepsilon_\alpha \left(c_1 d(\gamma_\alpha, \nu_0) + \sqrt{I(\gamma_\alpha)}\right)^2 + o(1). \tag{3.35}$$

As a corollary, if $(\alpha_n)_{n\in\mathbb{N}}$ is the sequence given by Proposition 3.2-(d), then we have

$$\begin{aligned} &\frac{\alpha_n^2}{2(1-\varepsilon_{\alpha_n})} d^2(\pi_{\alpha_n}, \rho_{\alpha_n}) - \frac{\alpha_n^2}{2(1+\varepsilon_{\alpha_n})} d^2(\mu_{\alpha_n}, \gamma_{\alpha_n}) \\ &\quad \leq \frac{\varepsilon_{\alpha_n}}{(1-\varepsilon_{\alpha_n})} I(\rho_{\alpha_n}) + \frac{\varepsilon_{\alpha_n}}{(1+\varepsilon_{\alpha_n})} I(\gamma_{\alpha_n}) + o(1). \end{aligned} \tag{3.36}$$

Proof. We begin by proving (3.34). First note that if $d(\rho_\alpha, \pi_\alpha) = 0$, there is nothing to prove. We thus only prove the first statement in the case that $d(\rho_\alpha, \pi_\alpha) > 0$. To do so, we define the auxiliary function $\tilde{\mathcal{G}}_\alpha(\cdot)$ by

$$\begin{aligned} \tilde{\mathcal{G}}_\alpha(\cdot) &= -(1 - \varepsilon_\alpha)\mathcal{G}_\alpha(\pi_\alpha, \mu_\alpha, \cdot, \gamma_\alpha) \\ &= \frac{\alpha}{2} d^2(\cdot, \pi_\alpha) + (1 - \varepsilon_\alpha)\frac{\alpha}{2} d^2(\cdot, \gamma_\alpha) + \varepsilon_\alpha \bar{\mathcal{E}}(\cdot) + c, \end{aligned}$$

where c is a constant. We obtain from (3.4), the definition of $\mathcal{B}_\alpha$ (see (3.5)) and the Lipschitzianity of Tataru's distance that

$$\forall \rho \in E, \quad \tilde{\mathcal{G}}_\alpha(\rho_\alpha) - \tilde{\mathcal{G}}_\alpha(\rho) \leq (1 - \varepsilon_\alpha)\alpha^{-1} d_T(\rho, \rho_\alpha) \leq (1 - \varepsilon_\alpha)\alpha^{-1} d(\rho, \rho_\alpha).$$

Let us now consider a geodesic $\boldsymbol{\zeta}^{\rho_\alpha \to \pi_\alpha}$, fix $\theta > 0$ small enough, and consider the curve $\boldsymbol{\zeta}_\theta^{\rho_\alpha \to \pi_\alpha}$ given by Assumption 2.9. Choosing $\rho = \boldsymbol{\zeta}_\theta^{\rho_\alpha \to \pi_\alpha}(s)$ in the above estimate and, dividing by s, and letting $s \downarrow 0$ we obtain

$$\liminf_{s\downarrow 0}\frac{\alpha}{2s}[d^2(\rho_\alpha,\pi_\alpha)-d^2(\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s),\pi_\alpha)] \tag{3.37a}$$

$$\leq \limsup_{s\downarrow 0}\frac{(1-\varepsilon_\alpha)}{\alpha s}d(\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s),\rho_\alpha) \tag{3.37b}$$

$$+\limsup_{s\downarrow 0}\frac{(1-\varepsilon_\alpha)\alpha}{2s}[d^2(\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s),\gamma_\alpha)-d^2(\rho_\alpha,\gamma_\alpha)] \tag{3.37c}$$

$$+\liminf_{s\downarrow 0}\frac{\varepsilon_\alpha}{s}[\bar{\mathcal{E}}(\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s))-\bar{\mathcal{E}}(\rho_\alpha)]. \tag{3.37d}$$

We start with estimates for all the terms on the right-hand side of (3.37). To this aim, we observe that for any $\sigma\in E$ we have, using the triangle inequality, the geodesic property and hypothesis (2.6)

$$\begin{aligned} d(\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s),\sigma) &\leq d(\rho_\alpha,\sigma)+d(\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s),\rho_\alpha)\\ &\leq d(\rho_\alpha,\sigma)+sd(\rho_\alpha,\pi_\alpha)+d(\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s),\boldsymbol{\zeta}^{\rho_\alpha\to\pi_\alpha}(s))\\ &\leq d(\rho_\alpha,\sigma)+sd(\rho_\alpha,\pi_\alpha)+s\theta(1+o(1)). \end{aligned} \tag{3.38}$$

Choosing $\sigma=\rho_\alpha$ to bound (3.37b), $\sigma=\gamma_\alpha$ for (3.37c), and $\sigma=\nu_0$ to bound the distance term of (3.37d) together with

$$\liminf_{s\downarrow 0}\frac{1}{s}[\mathcal{E}(\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s))-\mathcal{E}(\rho_\alpha)] \overset{(2.7)}{\leq} \sqrt{I(\rho_\alpha)}(d(\rho_\alpha,\pi_\alpha)+\theta) \tag{3.39}$$

for the energy term of (3.37d), we obtain that the right hand side in (3.37) is bounded above by

$$(d(\pi_\alpha,\rho_\alpha)+\theta)\left((1-\varepsilon_\alpha)(\alpha^{-1}+\alpha d(\rho_\alpha,\gamma_\alpha))+\varepsilon_\alpha\left(c_1 d(\rho_\alpha,\nu_0)+\sqrt{I(\rho_\alpha)}\right)\right). \tag{3.40}$$

Let us now turn the attention to (3.37a). Here, using that

$$\begin{aligned} d(\pi_\alpha,\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s)) &\leq (1-s)d(\rho_\alpha,\pi_\alpha)+d(\boldsymbol{\zeta}^{\rho_\alpha\to\pi_\alpha}(s),\boldsymbol{\zeta}_\theta^{\rho_\alpha\to\pi_\alpha}(s))\\ &\leq (1-s)d(\rho_\alpha,\pi_\alpha)+s\theta(1+o(1)) \end{aligned}$$

we find that (3.37a) is bounded below by

$$\alpha d(\rho_\alpha,\pi_\alpha)(d(\rho_\alpha,\pi_\alpha)-\theta). \tag{3.41}$$

Assembling together (3.40) with (3.41), dividing by $d(\rho_\alpha,\pi_\alpha)$ and letting $\theta\to 0$ yields

$$\alpha d(\rho_\alpha,\pi_\alpha)\leq\left((1-\varepsilon_\alpha)(\alpha^{-1}+\alpha d(\rho_\alpha,\gamma_\alpha))+\varepsilon_\alpha\left(c_1 d(\rho_\alpha,\nu_0)+\sqrt{I(\rho_\alpha)}\right)\right),$$

from which the bound (3.34) is obtained taking squares on both sides, using convexity of the square function on the right hand side and eventually dividing by two. Let us now proceed to the proof of the second inequality. We do the proof in detail as, even though it uses some arguments similar to those used to obtain the first estimate, there are also some non trivial differences. We begin by noting that we can assume without loss of generality that $d(\rho_\alpha, \gamma_\alpha) > 0$. Next, define the auxiliary test function $\bar{\mathcal{G}}_\alpha(\cdot)$ by

$$\begin{aligned}\bar{\mathcal{G}}_\alpha(\cdot) &= -\mathcal{G}_\alpha(\pi_\alpha, \mu_\alpha, \rho_\alpha, \cdot)\\ &= \frac{\alpha}{2(1+\varepsilon_\alpha)} d^2(\cdot, \mu_\alpha) + \frac{\alpha}{2} d^2(\rho_\alpha, \cdot) + \frac{\varepsilon_\alpha}{(1+\varepsilon_\alpha)} \bar{\mathcal{E}}(\cdot) + c.\end{aligned}$$

We obtain from (3.4), the definition of $\mathcal{B}_\alpha$ (see (3.5)) and the Lipschitzianity of Tataru's distance that

$$\forall \gamma \in E, \quad \bar{\mathcal{G}}_\alpha(\gamma_\alpha) - \bar{\mathcal{G}}_\alpha(\gamma) \leq \alpha^{-1} d_T(\gamma, \gamma_\alpha) \leq \alpha^{-1} d(\gamma, \gamma_\alpha).$$

Let us now consider a geodesic from γ_α to ρ_α, $\boldsymbol{\zeta}^{\gamma_\alpha \to \rho_\alpha}$, (Due to the fact that we don't have linearity and all the properties of the flow given in Assumption 2.9 are given with lim sup, we have to go from γ_α to ρ_α while for the other inequality we had to go from ρ_α to π_α) fix a $\theta > 0$ small enough, and consider the curve $\boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}$ given by Assumption 2.9. Using the previous estimate, we have, for all s small enough,

$$\liminf_{s \downarrow 0} \frac{\alpha}{2s} [d^2(\rho_\alpha, \gamma_\alpha) - d^2(\rho_\alpha, \boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}(s))] \tag{3.42a}$$

$$\leq \limsup_{s \downarrow 0} \frac{1}{\alpha s} d(\boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}(s), \gamma_\alpha) \tag{3.42b}$$

$$+ \limsup_{s \downarrow 0} \frac{\alpha}{2s(1+\varepsilon_\alpha)} \left(d^2(\boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}(s), \mu_\alpha) - d^2(\gamma_\alpha, \mu_\alpha)\right) \tag{3.42c}$$

$$+ \liminf_{s \downarrow 0} \frac{\varepsilon_\alpha}{s(1+\varepsilon_\alpha)} [\bar{\mathcal{E}}(\boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}(s)) - \bar{\mathcal{E}}(\gamma_\alpha)]. \tag{3.42d}$$

In order to estimate all terms containing d on the right hand side, we use the analogous of (3.38), namely that for all $\sigma \in E$

$$\begin{aligned}d(\boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}(s), \sigma) &\leq d(\gamma_\alpha, \sigma) + d(\boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}(s), \gamma_\alpha)\\ &\leq d(\gamma_\alpha, \sigma) + s d(\gamma_\alpha, \rho_\alpha) + d(\boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}(s), \boldsymbol{\zeta}^{\gamma_\alpha \to \rho_\alpha}(s))\\ &\leq d(\gamma_\alpha, \sigma) + s d(\gamma_\alpha, \rho_\alpha) + s\theta(1 + o(1)).\end{aligned} \tag{3.43}$$

Indeed, choosing $\sigma = \gamma_\alpha$ to bound the right hand side of (3.42b), $\sigma = \mu_\alpha$ to bound (3.42c), $\sigma = \nu_0$ to bound the distance term of (3.42d) and

$$\liminf_{s \downarrow 0} \frac{1}{s} [\mathcal{E}(\boldsymbol{\zeta}_\theta^{\gamma_\alpha \to \rho_\alpha}(s)) - \mathcal{E}(\gamma_\alpha)] \overset{(2.7)}{\leq} \sqrt{I(\gamma_\alpha)} (d(\gamma_\alpha, \rho_\alpha) + \theta) \tag{3.44}$$

for the energy term of (3.42d), we obtain that the right hand side in (3.42) is bounded above by

$$(d(\gamma_\alpha,\rho_\alpha)+\theta)\left(\alpha^{-1}+\frac{\alpha}{(1+\varepsilon_\alpha)}d(\gamma_\alpha,\mu_\alpha)+\frac{\varepsilon_\alpha}{(1+\varepsilon_\alpha)}\left(c_1 d(\gamma_\alpha,\nu_0)+\sqrt{I(\gamma_\alpha)}\right)\right) \quad (3.45)$$

Let us now turn the attention to (3.42a). Here, using that

$$\begin{aligned} d(\rho_\alpha,\boldsymbol{\zeta}_\theta^{\gamma_\alpha\to\rho_\alpha}(s)) &\le (1-s)d(\rho_\alpha,\gamma_\alpha)+d(\boldsymbol{\zeta}^{\gamma_\alpha\to\rho_\alpha}(s),\boldsymbol{\zeta}_\theta^{\gamma_\alpha\to\rho_\alpha}(s)) \\ &\le (1-s)d(\rho_\alpha,\gamma_\alpha)+s\theta(1+o(1)) \end{aligned}$$

we obtain that (3.42a) is bounded below by

$$\alpha d(\rho_\alpha,\gamma_\alpha)(d(\rho_\alpha,\gamma_\alpha)-\theta). \quad (3.46)$$

Assembling together (3.45) with (3.46), dividing by $d(\rho_\alpha,\gamma_\alpha)$ and letting $\theta\to 0$ yields

$$\alpha d(\rho_\alpha,\gamma_\alpha)-\alpha^{-1}\le\frac{\alpha}{(1+\varepsilon_\alpha)}d(\gamma_\alpha,\mu_\alpha)+\frac{\varepsilon_\alpha}{(1+\varepsilon_\alpha)}\left(c_1 d(\gamma_\alpha,\nu_0)+\sqrt{I(\gamma_\alpha)}\right)$$

If $\alpha d(\rho_\alpha,\gamma_\alpha)-\alpha^{-1}\ge 0$ the bound (3.35) is obtained taking squares on both sides, using convexity of the square function on the right hand side and the fact that $d(\rho_\alpha,\gamma_\alpha)$ is $o(1)$. If $\alpha d(\rho_\alpha,\gamma_\alpha)-\alpha^{-1}<0$, it is easily seen that the right hand side of (3.35) is bounded above by a function that is $o(1)$, from which the desired conclusion follows. Finally, the bound (3.36) is a consequence of (3.35), (3.34), Proposition 3.2-(d) and the basic inequality

$$\frac{1}{2}\left(c_1 d(\cdot,\nu_0)+\sqrt{I(\cdot)}\right)^2\le c_1^2 d^2(\cdot,\nu_0)+I(\cdot). \quad \square$$

4. Consequences of EVI and properties of the Tataru distances

4.1. Consequences of EVI

In this section we deduce from EVI various estimates on the behavior of d, $\mathcal{E}$ and I along the gradient flow. These estimates play a fundamental role in the proof of the comparison principle and are be obtained with little effort from those of [43].

Lemma 4.1. *Let Assumption 2.3 and 2.5 hold (in particular EVI inequality (EVI_κ)). For $\mu\in E$, let $(\mu(t))_{t\ge 0}$ be the corresponding gradient flow starting at μ. Then the following holds:*

(a) For each $c_1>-\kappa$ and for each $\nu\in E$ there exist $c_2,\tilde{c}_2\in\mathbb{R}$ such that if we set

$$\forall\pi\in E,\quad \bar{\mathcal{E}}(\pi):=\mathcal{E}(\pi)+\frac{c_1}{2}d^2(\pi,\nu)+c_2,$$

then we have

$$\inf_{\pi\in E} \bar{\mathcal{E}}(\pi) = 0 \tag{4.1}$$

and

$$\forall \pi \in E \quad \bar{\mathcal{E}}(\pi) \geq \frac{\kappa + c_1}{2} d^2(\pi, \nu) + \tilde{c}_2.$$

(b) For any $t > 0$ we have

$$\mathcal{E}(\mu(t)) - \mathcal{E}(\mu) = -\int_0^t I(\mu(s))\mathrm{d}s. \tag{4.2}$$

(c) The domain $\mathcal{D}(I)$ is dense in $\mathcal{D}(\mathcal{E})$ and dense in E. In particular, the domain $\mathcal{D}(\mathcal{E})$ of $\mathcal{E}$ is dense in E.
(d) For any $t > 0$, we have $I(\mu(t)) < \infty$. The map $t \mapsto I(\mu(t))$ is right-continuous at any $t_0 \geq 0$ such that $I(\mu(t_0)) < \infty$.
(e) Let $\nu \in E$ and let $(\nu(t))_{t\geq 0}$ be the corresponding gradient flow starting at ν. Then we have

$$d(\mu(t), \nu(t)) \leq e^{-\kappa t} d(\mu, \nu) \quad \forall t \in [0, +\infty). \tag{4.3}$$

In particular, for a given $\mu \in E$, there is at most one solution of (EVI_κ) such that $\mu(t) \to \mu$ as $t \to 0$.

Proof. We begin by observing that under the current hypothesis the triplet $(E, d, \mathcal{E})$ is a metric-functional system in the sense of [43, Eq 3.1]. This allows us to deduce most of the results we need to prove from Theorem 3.5 therein. Item (a) is proven at [43, Thm 3.5], see Eq (3.15) and the discussion surrounding its proof. For the proof of (b), note that [43, Thm 3.5, Eq 3.11] and [10, Theorem 2.1.7] imply that $t \mapsto \mathcal{E}(\mu(t))$ is locally Lipschitz and, hence, absolutely continuous on $(0, \infty)$. By [43, Thm 3.5, Eq 3.17] and the monotone convergence theorem, we obtain (4.2). Item (c) follows from [43, Thm 2.10] and Assumption 2.3, item (d) follows from [43, Thm 3.5, Eq 3.11 and Eq 3.12] and item (e) from [43, Thm 3.5, Eq 3.10]. □

4.2. Properties of the Tataru distance

We develop here the key results that hold for our adjusted Tataru distance. First of all, note that the infimum in the definition is attained.

Remark 4.2. Since the gradient flow (thanks to Assumption 2.5) and the distance d are continuous then the inf is attained. Indeed, for all $\mu, \nu \in E$, we have $0 \leq d_T(\mu, \nu) \leq 0 + d(\mu, \nu(0)) = d(\mu, \nu)$. Let $(t_n)_{n\in\mathbb{N}} \in [0, +\infty)$ be a minimizing sequence, i.e.

$$\lim_{n\to+\infty} t_n + e^{\hat{\kappa}t_n} d(\mu,\nu(t_n)) = d_T(\mu,\nu).$$

Then, for all $n \in \mathbb{N}$ we have

$$0 \le t_n + e^{\hat{\kappa}t_n} d(\mu,\nu(t_n)) \le d(\mu,\nu),$$

hence $0 \le t_n \le d(\mu,\nu)$ and $(t_n)_{n\in\mathbb{N}}$ is a bounded sequence. Passing to a subsequence, still called $(t_n)_{n\in\mathbb{N}}$ by an abuse of notation, we have $\lim_{n\to+\infty} t_n = \bar{t}$ for a $\bar{t} \ge 0$. Being the gradient flow $\nu(\cdot)$ and d continuous we also have

$$\lim_{n\to+\infty} e^{\hat{\kappa}t_n} d(\mu,\nu(t_n)) = e^{\hat{\kappa}\bar{t}} d(\mu,\nu(\bar{t})).$$

Therefore we must have

$$d_T(\mu,\nu) = \bar{t} + e^{\hat{\kappa}\bar{t}} d(\mu,\nu(\bar{t})).$$

Secondly, we note that the EVI inequality (EVI_κ) leads to the control on the growth of the distance along two solutions of the gradient flow.

Lemma 4.3. *We have for all $\mu, \hat{\mu}, \nu, \hat{\nu} \in E$ and $r > 0$ that*

(a)

$$d_T(\mu,\nu) - d_T(\hat{\mu},\hat{\nu}) \le d(\mu,\hat{\mu}) + d(\nu,\hat{\nu})$$

(b)

$$\frac{d_T(\nu(r),\hat{\nu}) - d_T(\nu,\hat{\nu})}{r} \le 1.$$

Proof. For (a) Let $t \in [0,+\infty)$ be optimal for $d_T(\hat{\mu},\hat{\nu})$, i.e.

$$d_T(\hat{\mu},\hat{\nu}) = t + e^{\hat{\kappa}t} d(\hat{\mu},\hat{\nu}(t)).$$

Then, we have

$$\begin{aligned} d_T(\mu,\nu) - d_T(\hat{\mu},\hat{\nu}) &\le e^{\hat{\kappa}t} d(\mu,\nu(t)) - e^{\hat{\kappa}t} d(\hat{\mu},\hat{\nu}(t)) \\ &\le e^{\hat{\kappa}t} \left[d(\mu,\hat{\mu}) + d(\hat{\mu},\nu(t)) - d(\hat{\mu},\hat{\nu}(t)) \right] \\ &\le e^{\hat{\kappa}t} d(\mu,\hat{\mu}) + e^{\hat{\kappa}t} d(\nu(t),\hat{\nu}(t)) \\ &\le e^{\hat{\kappa}t} d(\mu,\hat{\mu}) + e^{(\hat{\kappa}-\kappa)t} d(\nu,\hat{\nu}) \\ &\le d(\mu,\hat{\mu}) + d(\nu,\hat{\nu}), \end{aligned}$$

where in line 4 we use equation (4.3), in line 5 we use that $\hat{\kappa} \le 0$ and $\hat{\kappa} - \kappa \le 0$. For (b), let $t \in [0,+\infty)$ be optimal for $d_T(\nu,\hat{\nu})$. Then working with the sub-optimal $t + r$ for the first term, we obtain

$$\begin{aligned}\frac{d_T(\nu(r),\hat\nu)-d_T(\nu,\hat\nu)}{r} &\le \frac{e^{(t+r)\hat\kappa}d(\nu(r),\hat\nu(t+r))+t+r-e^{t\hat\kappa}d(\nu,\hat\nu(t))-t}{r}\\ &\le \frac{e^{(t+r)\hat\kappa}d(\nu(r),\hat\nu(t+r))-e^{\hat\kappa t}d(\nu,\hat\nu(t))}{r}+1\\ &\le e^{\hat\kappa t}\frac{d(\nu,\hat\nu(t))-d(\nu,\hat\nu(t))}{r}+1\\ &\le 1\end{aligned}$$

by equation (4.3) and the fact that $\hat\kappa-\kappa\le 0$, $\hat\kappa\le 0$. □

Lemma 4.4. *For $\rho,\mu,\nu\in E$, we have*

$$d_T(\rho,\nu)\le d_T(\rho,\mu)+d_T(\mu,\nu).$$

Proof. We have

$$\begin{aligned}d_T(\rho,\nu) &= \inf_{t\ge 0}\left\{t+e^{\hat\kappa t}d(\rho,\nu(t))\right\}\\ &= \inf_{t,s\ge 0}\left\{t+s+e^{\hat\kappa(t+s)}d(\rho,\nu(t+s))\right\}\\ &\le \inf_{t,s\ge 0}\left\{t+s+e^{\hat\kappa(t+s)}d(\rho,\mu(t))+e^{\hat\kappa(t+s)}d(\mu(t),\nu(t+s))\right\}.\end{aligned}$$

We now use that, as $\hat\kappa\le 0$ we have $e^{\hat\kappa(t+s)}d(\rho,\mu(t))\le e^{\hat\kappa t}d(\rho,\mu(t))$. For the term $e^{\hat\kappa(t+s)}d(\mu(t),\nu(t+s))$ we use equation (4.3) and the fact that $\hat\kappa-\kappa\le 0$. This yields

$$\begin{aligned}d_T(\rho,\nu) &\le \inf_{t,s\ge 0}\left\{t+s+e^{\hat\kappa t}d(\rho,\mu(t))+e^{\hat\kappa s}d(\mu,\nu(s))\right\}\\ &\le d_T(\rho,\mu)+d_T(\mu,\nu). \quad \square\end{aligned}$$

5. Examples

In this section, we treat three key examples:

- Hilbert spaces, in particular in the context where $\mathcal{E}$ is derived from a Dirichlet energy. This includes e.g. the linearly controlled Allen-Cahn equation.
- Finite dimensional spaces that are essentially Riemannian manifolds.
- The Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$.

In all the examples, the first step is the verification that the metric space satisfies Assumption 2.3 and that there exists a gradient flow satisfying (EVI_κ). We will argue this final point starting from κ-convexity of the functional $\mathcal{E}$, see Definition 5.1 below. In concrete examples, this property is typically easier to verify, and is strongly related to

(EVI_κ). Indeed, κ-convexity of $\mathcal{E}$ is implied by the existence of a gradient flow satisfying (EVI_κ) by a result of [24]. The other implication is not established in general, but includes an extensive list of relevant examples, see the discussion in Section 3.4 of [43]. For our first two examples we will argue via this route, while for the final example, we will use the methods of [3] based on the κ-convexity of $\mathcal{E}$ along generalized geodesics.

Definition 5.1. Let $\kappa \in \mathbb{R}$. We say that a lower semi-continuous functional $\mathcal{E} : E \to \mathbb{R} \cup \{\infty\}$ is κ-convex on a curve $\gamma : [0,1] \to \mathcal{D}(\mathcal{E})$ if it satisfies for all $t \in [0,1]$ the inequality

$$\mathcal{E}\left(\gamma(t)\right) \leq (1-t)\mathcal{E}(\gamma(0))) + t\mathcal{E}(\gamma(1))) - \frac{\kappa}{2} t(1-t) d^2(\gamma(0), \gamma(1)).$$

If for any two points $\rho, \pi \in \mathcal{D}(\mathcal{E})$, there exists a constant speed geodesic $\boldsymbol{\zeta}^{\rho\to\pi}$ such that $\mathcal{E}$ is κ-convex on $\boldsymbol{\zeta}^{\rho\to\pi}$, then we call $\mathcal{E}$ κ-convex. If $\mathcal{E}$ is κ-convex on all geodesics, then we call $\mathcal{E}$ strongly κ-convex.

Theorem 5.2 *(Theorem 3.2 [24]). Consider a lower semi-continuous functional $\mathcal{E} : E \to \mathbb{R}\cup\{\infty\}$ on a geodesic space (E,d) such that there exist a gradient flow satisfying (EVI_κ). Then $\mathcal{E}$ is strongly κ-convex.*

Therefore, in all examples below, we can outright assume that we are working with a κ-convex functional. In this context, the following proposition simplifies establishing Assumption 2.9.

Proposition 5.3. *Consider the context of Assumption 2.3. Consider ρ, π such that $I(\rho) + \mathcal{E}(\pi) < \infty$ and let $\boldsymbol{\zeta}^{\rho\to\pi}$ be the constant speed geodesic between ρ and π. Suppose that for each $\theta > 0$ there is a curve $(\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t))_{t\in[0,1]}$, $\boldsymbol{\zeta}_\theta^{\rho\to\pi}(0) = \rho$, $\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t) \neq \rho$ if $t \in [0,1]$ such that:*

(a) $\mathcal{E}$ is κ-convex along $\boldsymbol{\zeta}_\theta^{\rho\to\pi}$,
(b) the angle condition (2.6) holds:

$$\limsup_{t\downarrow 0} \frac{d(\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t), \boldsymbol{\zeta}^{\rho\to\pi}(t))}{t} \leq \theta,$$

(c) for all t we have $\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t) \in \mathcal{D}(|\partial\mathcal{E}|)$ and

$$\lim_{t\downarrow 0} |\partial\mathcal{E}|(\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t)) = |\partial\mathcal{E}|(\rho).$$

Then Assumption 2.9 holds.

Remark 5.4. Consider the context in which the approximating curves $\boldsymbol{\zeta}_\theta^{\rho\to\pi}$ are themselves geodesics. Then by Theorem 5.2 we obtain that $\mathcal{E}$ is κ-convex along geodesics implying (a).

Remark 5.5. In a range of contexts, one finds that $I = |\partial\mathcal{E}|^2$ is convex along geodesics inside $\mathcal{D}(|\partial\mathcal{E}|)$. As $|\partial\mathcal{E}|$ is always lower semi-continuous, this implies (c).

Proof. By assumption (b), it suffices the establish (2.7) for the curves $\zeta_\theta^{\rho\to\pi}$. Due to the κ-convexity of $\mathcal{E}$ along $\zeta_\theta^{\rho\to\pi}$ given in (a), we can apply Proposition 2.4.9 in [3] to obtain

$$d(\zeta_\theta^{\rho\to\pi}(t),\rho)|\partial\mathcal{E}(\zeta_\theta^{\rho\to\pi}(t))| \geq \mathcal{E}(\zeta_\theta^{\rho\to\pi}(t)) - \mathcal{E}(\rho) + \frac{\kappa}{2}d^2(\zeta_\theta^{\rho\to\pi}(t),\rho).$$

Rewriting the inequality yields

$$\frac{\mathcal{E}(\zeta_\theta^{\rho\to\pi}(t)) - \mathcal{E}(\rho)}{t} \leq \frac{d(\zeta_\theta^{\rho\to\pi}(t),\rho)}{t}|\partial\mathcal{E}(\zeta_\theta^{\rho\to\pi}(t))| - \frac{\kappa}{2t}d^2(\zeta_\theta^{\rho\to\pi}(t),\rho).$$

Using the triangle inequality, and the angle condition of (b), and that $\zeta^{\rho\to\pi}$ is a geodesic, we find

$$\limsup_{t\downarrow 0}\frac{d(\zeta_\theta^{\rho\to\pi}(t),\rho)}{t} \leq \limsup_{t\downarrow 0}\frac{d(\zeta_\theta^{\rho\to\pi}(t),\zeta^{\rho\to\pi}(t)) + d(\zeta^{\rho\to\pi}(t),\rho)}{t} \leq \theta + d(\rho,\pi).$$

Combining the two above equations, we have

$$\begin{aligned}\liminf_{t\downarrow 0}\frac{\mathcal{E}(\zeta_\theta^{\rho\to\pi}(t)) - \mathcal{E}(\rho)}{t} &\leq (\theta + d(\rho,\pi))\liminf_{t\downarrow 0}|\partial\mathcal{E}(\zeta_\theta^{\rho\to\pi}(t))| \\ &\leq (\theta + d(\rho,\pi))\liminf_{t\downarrow 0}|\partial\mathcal{E}(\rho)|\end{aligned}$$

establishing the claim. □

5.1. Hilbert spaces

In this subsection, we assume that $(E,d) = (\mathcal{H},\|\cdot\|)$ is a Hilbert space. Below we will verify our Assumptions in two examples, one treats linearly controlled Ornstein-Uhlenbeck type Hamiltonians on general Hilbert spaces, the other treats $L^2(\mathbb{R}^d)$ with an energy that yields the solution to the Allen-Cahn equation as a gradient flow. For another example where our methods apply see [30]. We start out with a general existence result for (EVI_κ).

Theorem 5.6 *(Brezis-Pazy, Theorem 3.1 [2]). Let $\mathcal{E}$ be κ-convex and lower semi-continuous. Then there is a unique solution to (EVI_κ) for $\mathcal{E}$.*

5.1.1. The gradient flow constructed from a maximally dissipative operator

As the main example representing a large class of flows, we consider

$$\dot{\rho} = \frac{1}{2}\Delta\rho - \kappa\rho \tag{5.1}$$

on $L^2(\mathbb{R}^d)$ which formally corresponds to the gradient flow of

$$\mathcal{E}(\rho) = \frac{1}{2}\int |\nabla\rho(x)|^2 + \kappa|\rho(x)|^2 \mathrm{d}x = -\frac{1}{2}\langle \Delta\rho - \kappa\rho, \rho\rangle. \tag{5.2}$$

We see that $\mathcal{E}$ decomposes as a Dirichlet energy which is lower semi-continuous and convex, combined with $\kappa/2$ times the norm-squared. This implies $\mathcal{E}$ is κ-convex and that the gradient flow satisfying (EVI_κ) represented by (5.1) exists by Theorem 5.6. The use of the Laplacian or the specific form of the Hilbert space in this argument is not essential. The example thus generalizes immediately to the context where we consider a general Hilbert space $\mathcal{H}$ and replace Δ in (5.1) by a maximally dissipative *linear* self-adjoint operator C. We introduce some definitions to take care of general maximally dissipative operators and establish their connection 0-convex energy functionals.

Definition 5.7. We say that an operator $C \subseteq E \times E$ is dissipative, if for all $(\rho_1, \xi_1), (\rho_2, \xi_2) \in C$ we have

$$\langle \xi_1 - \xi_2, \rho_1 - \rho_2\rangle \leq 0.$$

If C is a single-valued operator, dissipativity is equivalent to

$$\langle C\rho_1 - C\rho_2, \rho_1 - \rho_2\rangle \leq 0$$

for all $\rho_1, \rho_2 \in \mathcal{D}(C)$. We say that an operator C is maximally dissipative if any dissipative extension B of the operator C equals C.

In the context of a maximally dissipative linear and self-adjoint operator, which include all self-adjoint generators of linear strongly continuous semigroups, we thus identify the flow of this semigroup as the gradient flow for the Dirichlet energy constructed from C.

Proposition 5.8. *Let $(C, \mathcal{D}(C))$ be a at most single-valued linear self-adjoint and maximally dissipative operator on $\mathcal{H}$ and let $\kappa \in \mathbb{R}$. Let $\mathcal{E}$ be the lower semi-continuous regularization of the functional*

$$\mathcal{E}_0(\rho) := \begin{cases} -\frac{1}{2}\langle C\rho, \rho\rangle + \frac{\kappa}{2}\|\rho\|^2 & \textit{if } \rho \in \mathcal{D}(C), \\ \infty & \textit{otherwise} \end{cases}$$

Then the conclusion of Theorem 2.14 hold for the Hilbert space $\mathcal{H}$ and energy functional $\mathcal{E}$.

For the proof, we turn to Theorem 2.14 and verify Assumptions 2.3, 2.5 and 2.9. As the first assumption is immediate in this Hilbertian context, we focus on the other two assumptions. To facilitate the verification, we first study the convexity properties and the Frechét subdifferential of $\mathcal{E}_0$ and $\mathcal{E}$ in the case that $\kappa = 0$.

Definition 5.9. Let $\phi : E \to \mathbb{R} \cup \{\infty\}$ be a functional. The Frechét subdifferential $\partial\phi(x)$ at $x \in E$ is given by

$$\partial\phi(\rho) := \left\{ \xi \in E \,\middle|\, \liminf_{\pi\to\rho} \frac{\phi(\pi) - \phi(\rho) - \langle \xi, \pi - \rho \rangle}{\|\pi - \rho\|} \geq 0 \right\}. \tag{5.3}$$

If ϕ is lower semi-continuous and convex then by Proposition 1.4.4 of [2] also

$$\partial\phi(\rho) = \{\xi \in E \,|\, \forall\, \pi \in E\colon\ \phi(\pi) - \phi(\rho) - \langle \xi, \pi - \rho \rangle \geq 0\}. \tag{5.4}$$

Note that the notation $|\partial\phi|(\rho)$ for the local slope of ϕ at ρ should not be interpreted as the ‘size’ of $\partial\phi(\rho)$, although the local slope is related to the size of the smallest element in $\partial\phi(\rho)$. See Proposition 1.4.4 of [2].

Lemma 5.10. *Consider the setting of Proposition 5.8 with $\kappa = 0$. We then have that*

(a) $\mathcal{E}_0 \geq 0$ and for $\rho, \pi \in \mathcal{D}(C)$ and $t \in [0,1]$ we have

$$\mathcal{E}_0(\pi) - \mathcal{E}_0(\rho) - \langle -C\rho, \pi - \rho \rangle = \mathcal{E}_0(\pi - \rho) \geq 0, \tag{5.5}$$

$$\mathcal{E}_0(\rho + t(\pi - \rho)) = (1-t)\mathcal{E}_0(\rho) + t\mathcal{E}_0(\pi) - t(1-t)\mathcal{E}_0(\pi - \rho). \tag{5.6}$$

(b) $\mathcal{D}(C) \subseteq \mathcal{D}(\mathcal{E})$, $0 \leq \mathcal{E} \leq \mathcal{E}_0$ and $\mathcal{E} = \mathcal{E}_0$ on $\mathcal{D}(C)$ and $\mathcal{E}$ is 0-convex. If ρ is such that $\mathcal{E}(\rho) < \infty$ then there are $\rho_n \in \mathcal{D}(C)$ such that

$$\lim_n \mathcal{E}_0(\rho_n) = \lim_n \mathcal{E}(\rho_n) = \mathcal{E}(\rho). \tag{5.7}$$

(c) $\mathcal{D}(\partial\mathcal{E}) = \mathcal{D}(C)$ and for all $\rho \in \mathcal{D}(C)$ we have $\partial\mathcal{E}(\rho) = \{-C\rho\}$ and $|\partial\mathcal{E}|(\rho) = \|C\rho\|$.

Proof. For the proof of (a), note that due to dissipativity $\mathcal{E}_0 \geq 0$. Next, consider $x, y \in \mathcal{D}(C)$, then using the linearity of C we obtain

$$\mathcal{E}_0(\pi) - \mathcal{E}_0(\rho) - \langle -C\rho, \pi - \rho \rangle = \mathcal{E}_0(\pi - \rho) + \frac{1}{2}\left(\langle C\rho, \pi \rangle - \langle C\pi, \rho \rangle\right).$$

As C is self-adjoint, we have

$$\mathcal{E}_0(\pi) - \mathcal{E}_0(\rho) - \langle -C\pi, \pi - \rho \rangle = \mathcal{E}_0(\pi - \rho) \geq 0$$

establishing (5.5). The parallelogram rule in (5.6) follows by a direct computation. We proceed to the second item. As $\mathcal{E}$ is the lower semi continuous regularization of $\mathcal{E}_0 \geq 0$, we find $0 \leq \mathcal{E} \leq \mathcal{E}_0$. Thus, let $\rho \in \mathcal{D}(C)$ and consider $\rho_n \in \mathcal{D}(C)$ such that $\rho_n \to \rho$. Then by (a), we have

$$\liminf_{n\to\infty} \mathcal{E}_0(\rho_n) \geq \liminf_{n\to\infty} \mathcal{E}_0(\rho) + \langle -C\rho, \rho_n - \rho\rangle = \mathcal{E}_0(\rho)$$

establishing that $\mathcal{E}(\rho) = \mathcal{E}_0(\rho)$. As a consequence, the 0-convexity of $\mathcal{E}$ follows from (5.6). (5.7) follows by construction. To establish (c), first consider $\rho \in \mathcal{D}(C)$. We verify that $-C\rho \in \partial\mathcal{E}(\rho)$ by using (5.4), in other words, we establish

$$\mathcal{E}(\pi) - \mathcal{E}(\pi) - \langle -C\rho, \pi - \rho\rangle \geq 0$$

for any π. First note that if $\mathcal{E}(\pi) = \infty$ there is nothing to prove. So consider π such that $\mathcal{E}(\pi) < \infty$. By (5.7) there are $\pi_n \in \mathcal{D}(C)$ converging to π satisfying $\lim_n \mathcal{E}(\pi_n) = \mathcal{E}(\pi)$. Then by (5.5) we have

$$\begin{aligned}&\mathcal{E}(\pi) - \mathcal{E}(\rho) - \langle C\rho, \pi - \rho\rangle \\ &\qquad \geq \lim_n \mathcal{E}(\pi_n) - \mathcal{E}(\rho) - \langle -C\rho, \pi_n - \rho\rangle \geq \lim_n \mathcal{E}(\pi_n - \rho) \geq 0\end{aligned}$$

so that $\rho \in \mathcal{D}(\partial\mathcal{E})$ and $-C\rho \in \partial\mathcal{E}(\rho)$. It follows that the graph of C is contained in the dissipative operator $-\partial\mathcal{E}$ and as C is maximally dissipative $C = -\partial\mathcal{E}$. We thus find that $\partial\mathcal{E}(\rho) = \{-C\rho\}$ which implies by Proposition 1.4.4 of [3] that $|\partial\mathcal{E}|(\rho) = \|C\rho\|$. □

Proof of Proposition 5.8. It suffices to verify Assumptions 2.3, 2.5 and 2.9. First note that Assumption 2.3 is immediate. We next turn to Assumption 2.5 and establish the existence of the gradient flow satisfying (EVI_κ). As the map $\rho \mapsto \frac{\kappa}{2}\|\rho\|^2$ is κ-convex, it follows by Lemma 5.10 that $\mathcal{E}$ is κ-convex. Thus, Theorem 5.6 implies the existence of a solution to (EVI_κ) establishing Assumption 2.5. We will verify Assumption 2.9 by means of Proposition 5.3. Consider ρ, π such that $I(\rho)+\mathcal{E}(\pi) < \infty$. We approximate the geodesic $\boldsymbol{\zeta}^{\rho\to\pi}(t) = (1-t)\rho + t\pi$ between ρ and π by the geodesic $\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t) := (1-t)\rho + tS(\theta)\pi$ between ρ and $S(\hat{\theta})\pi$, where $t \mapsto S(t)\pi$ is used to denote the gradient flow started from π and where $\hat{\theta}$ is chosen such that $\left\|S(\hat{\theta})\pi - \pi\right\| \leq \theta$. To verify the angle condition (b) of Proposition 5.3, note that

$$\begin{aligned}\frac{\|\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t) - \boldsymbol{\zeta}^{\rho\to\pi}(t)\|}{t} &= \frac{\left\|(1-t)\rho + tS(\hat{\theta})\pi - ((1-t)\rho + t\pi)\right\|}{t} \\ &= \left\|S(\hat{\theta})\pi - \pi\right\|\end{aligned}$$

which by choice of $\hat{\theta}$ is smaller than θ. For the second property, note that by Lemma 5.10 we have

$$\lim_{t\downarrow 0} |\partial\mathcal{E}|(\boldsymbol{\zeta}_\theta^{\rho\to\pi}(t)) = \lim_{t\downarrow 0} \left\|(1-t)C\rho + tCS(\hat{\theta})\pi\right\| = \|C\rho\| = |\partial\mathcal{E}|(\rho)$$

so that (2.7) follows by Proposition 5.3. □

5.1.2. The Allen-Cahn equation

In the context of more concrete Hilbert spaces, we can introduce more general energy functionals. We will not aim for an exhaustive list, but rather consider a single example of interest: the energy functional associated to the Allen-Cahn equation on $\mathcal{H} = L^2(\mathbb{R}^d)$:

$$\dot{\rho} = \frac{1}{2}\Delta\rho - F'(\rho) - \kappa\rho. \tag{5.8}$$

Here $\kappa \in \mathbb{R}$ and $F : \mathbb{R}^d \to [0, \infty)$ is a non-negative convex C^1 function such that $F(0) = 0$. By Remark 2.3.9 and Corollary 1.4.5 in [3], we can represent this equation as the gradient flow of the energy

$$\mathcal{E}(\rho) = \frac{1}{2}\int |\nabla\rho(x)|^2 + \kappa|\rho(x)|^2 \mathrm{d}x + \int F(\rho(x))\mathrm{d}x. \tag{5.9}$$

Proposition 5.11. *Consider the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^d)$ and energy functional $\mathcal{E}$ of (5.9), where $\kappa \in \mathbb{R}$ and where $F : \mathbb{R}^d \to [0, \infty)$ is a non-negative convex C^1 function such that $F(0) = 0$. Then the conclusion of Theorem 2.14 hold.*

Proof. It suffices to verify Assumptions 2.3, 2.5 and 2.9. By construction, $\mathcal{E}$ is κ-convex. By Theorem 5.6 the gradient flow for $\mathcal{E}$ exists and satisfies (EVI_κ). As in the proof of Proposition 5.8, it thus suffices to establish Assumption 2.9. We do so as above. First note that by (3.4.14) of Remark 2.3.9 and Corollary 1.4.5 in [3] we have

$$\partial\mathcal{E}(\rho) = \begin{cases} \{\Delta\rho - F'(\rho) - \kappa\rho\} & \text{if } \Delta\rho, F'(\rho) \in L^2(\mathbb{R}^d), \\ \emptyset & \text{otherwise.} \end{cases}$$

We thus obtain that

$$|\partial\mathcal{E}|(\rho) = \|\Delta\rho - F'(\rho) - \kappa\rho\|.$$

We next establish the conditions for Proposition 5.3, and we do so on the basis of the same curves $\zeta_\theta^{\rho\to\pi}(t) = (1-t)\rho + tS(\hat{\theta})\pi$ as in the proof of Proposition 5.8. $\zeta_\theta^{\rho\to\pi}$ is therefore the linear interpolation between two elements in $\mathcal{D}(|\partial\mathcal{E})|$. As F' is increasing and Δ is linear, it follows that $\zeta_\theta^{\rho\to\pi}(t) \in \mathcal{D}(|\partial\mathcal{E}|)$ for all $t \in [0, 1]$. We next establish that

$$\lim_{t\downarrow 0} |\partial\mathcal{E}|^2(\zeta_\theta^{\rho\to\pi}(t)) = |\partial\mathcal{E}|^2(\rho). \tag{5.10}$$

We will establish this result by the use of the dominated convergence theorem. First of all

$$|\partial\mathcal{E}|^2(\zeta_\theta^{\rho\to\pi}(t)) = \int |\Delta\zeta_\theta^{\rho\to\pi}(t)(x) - F'(\zeta_\theta^{\rho\to\pi}(t)(x)) - \kappa\zeta_\theta^{\rho\to\pi}(t)(x)|^2 \mathrm{d}x$$

and as $\zeta_\theta^{\rho\to\pi}(t) \to \rho$ point-wise as $t \downarrow 0$, it suffices to find a integrable dominating function. Elementary point-wise estimates yield

$$\begin{aligned}
&|\Delta\zeta_\theta^{\rho\to\pi}(t)(x) - F'(\zeta_\theta^{\rho\to\pi}(t)(x)) - \kappa\zeta_\theta^{\rho\to\pi}(t)(x)|^2 \\
&\quad \leq 3|\Delta\zeta_\theta^{\rho\to\pi}(t)|^2 + 3|F'(\zeta_\theta^{\rho\to\pi}(t)(x)|^2 + 3\kappa^2|\zeta_\theta^{\rho\to\pi}(t)(x)|^2 \\
&\quad \leq 3|\Delta\rho(x)|^2 + 3|\Delta S(\hat{\theta})\pi(x)|^2 + 3|F'(\rho(x))|^2 \\
&\qquad + 3|F'(S(\hat{\theta})\pi(x))| + 3\kappa^2|\rho(x)|^2 + 3\kappa^2|S(\hat{\theta})\pi(x)|^2
\end{aligned}$$

as F' is increasing, and all six terms are integrable by assumption. Thus (5.10) follows by dominated convergence. Thus Assumption 2.9 follows by an application of Proposition 5.3. □

5.2. Almost Riemannian manifolds

In our second set of examples, we consider spaces that are essentially Riemannian manifolds. To illustrate what we are aiming for, consider the Hamiltonian

$$Hf(x) = (\mu - x)f'(x) + \frac{1}{2}x(f'(x))^2, \qquad x \geq 0 \tag{5.11}$$

for some constant $\mu > 0$. This Hamiltonian arises in the study of Freidlin-Wentzell type large deviation analysis of the Cox-Ingersoll-Ross model in finance [23,25]. Following [25], we study the Hamilton–Jacobi equation using a Riemannian point of view, where the Riemannian metric is generated by the quadratic part of the Hamiltonian. Arguing that the Hamiltonian is a map on the co-tangent bundle, we obtain a metric on the tangent bundle that satisfies $\langle v, w\rangle_{g(x)} = x^{-1}vw$ with the metric $g(x) = x^{-1}$ being singular in 0. We will show, however, that by interpreting the drift in (5.11) as the gradient flow of a functional $\mathcal{E}$ that satisfies $\mathcal{E}(0) = \infty$, we can work around the singularity of the metric at the boundary. The framework that we will be working in is the following.

Assumption 5.12. Let $(E, d, \mathcal{E})$ be a triple of a complete space (E, d) together with an energy $\mathcal{E} : E \to (-\infty, \infty]$. Assume that the following are satisfied.

(a) $E_0 := \mathcal{D}(\mathcal{E})$ is dense in E and the restriction of d to E_0 is such that (E_0, d) is a smooth Riemannian manifold.
(b) $\mathcal{E}$ is continuously differentiable on E_0.
(c) $\mathcal{E}$ is κ-convex along geodesics in E_0.

Proposition 5.13. *Suppose that Assumption 5.12 is satisfied, then the conclusion of Theorem 2.14 hold.*

Before giving the proof, we start with an auxiliary result that relates the slope to directional derivatives.

Definition 5.14. Let ϕ be a lower semi-continuous functional. Suppose $x \in \mathcal{D}(\phi)$. For a geodesic $\boldsymbol{\zeta}^{x\to y}$ denote the directional derivative of ϕ along the geodesic $\boldsymbol{\zeta}^{x\to y}$ by

$$\phi'(x, \boldsymbol{\zeta}^{x\to y}) := \liminf_{t\downarrow 0} \frac{\phi(\boldsymbol{\zeta}^{x\to y}(t)) - \phi(x)}{t}.$$

Lemma 5.15. *If ϕ is κ-convex on geodesics, then*

$$|\partial\phi|(x) = \|\operatorname{grad}\phi(x)\|_{T_x E_0},$$

and Assumption 2.9 is satisfied for ϕ.

Proof. For any two points $x, y \in \mathcal{D}(\mathcal{E})$ we will derive (2.6) and (2.7) with $\theta = 0$ for a geodesic $\boldsymbol{\zeta}^{x\to y}$. Using the κ-convexity of ϕ on geodesics, we derive as in [43, Section 2.3] that

$$|\partial\phi|(x) := \sup_{y\in\mathcal{D}(\phi),\ \text{geodesics}\ \boldsymbol{\zeta}^{x\to y}} \frac{\phi'(x, \boldsymbol{\zeta}^{x\to y})}{d(x,y)}.$$

As $\mathcal{E}$ is continuously differentiable on the domain of $\mathcal{E}$, we can obtain an upper bound on the directional derivative by using the Cauchy-Schwarz inequality

$$\begin{aligned}\phi'(x, \boldsymbol{\zeta}^{x\to y}) &= \lim_{t\downarrow 0} \frac{\phi(\boldsymbol{\zeta}^{x\to y}(t)) - \phi(x)}{t} = \langle \operatorname{grad}\phi(x), \dot{\boldsymbol{\zeta}}^{x\to y}(0)\rangle \\ &\le \|\operatorname{grad}\phi(x)\|_{T_x E_0} \left\|\dot{\boldsymbol{\zeta}}^{x\to y}(0)\right\|_{T_x E_0}.\end{aligned}$$

As $\boldsymbol{\zeta}^{x\to y}$ is a length-minimizing geodesic, we have $\left\|\dot{\boldsymbol{\zeta}}^{x\to y}(0)\right\|_{T_x E_0} = d(x,y)$, so that

$$|\partial\phi|(x) \le \|\operatorname{grad}\phi(x)\|_{T_x E_0}.$$

To establish the converse inequality, recall that on a Riemannian manifold geodesics are locally length minimizing. Thus there is some $\delta > 0$ such that the geodesic (in the Riemannian sense of the word) $\gamma : [0,1] \to E_0$ started at x in the direction $\operatorname{grad}\phi(x)$ of length δ satisfies $d(\gamma(0), \gamma(1)) = \delta$, and is thus a geodesic in our sense of the word. A direct computation yields that

$$\dot\gamma(0) = \frac{\delta}{\|\operatorname{grad}\phi(x)\|_{T_x E_0}} \operatorname{grad}\phi(x)$$

which implies

$$\phi'(x, \boldsymbol{\zeta}^{x\to y}) = \lim_{t\downarrow 0} \frac{\phi(\gamma(t)) - \phi(x)}{t} = \langle \operatorname{grad}\phi(x), \dot\gamma(0)\rangle = \delta \left\|\operatorname{grad}\mathcal{E}(x)\right\|_{T_x E_0}.$$

We can conclude that $|\partial\phi|(x) \leq \|\operatorname{grad}\phi(x)\|_{T_x E_0}$. For the proof of Assumption 2.9, we can take for all x, y and θ the geodesic $\boldsymbol{\zeta}^{x\to y}$ so that (2.6) is satisfied. Note that (2.7) can be verified using Cauchy-Schwarz as in the first part of this proof. □

Proof of Proposition 5.13. It suffices to verify Assumptions 2.3, 2.5 and 2.9. Assumption 2.3 is immediate. The gradient flow for $\mathcal{E}$ can be constructed by local arguments and by construction it remains in $\mathcal{D}(\mathcal{E})$. Assumption 2.5, or in other words, that the gradient flow satisfies (EVI_κ), follows by Proposition 23.1 in [49]. Assumption 2.9 follows from Lemma 5.15. □

For completeness, we verify the assumptions corresponding to the Hamiltonian of (5.11).

Lemma 5.16. *Assumption 5.12 is satisfied for* $E = \mathbb{R}^+$, $\mathcal{E}(x) = -\mu\log(x) + x - (\mu - \mu\log\mu)$ *and* $d(x, y) = 2|\sqrt{x} - \sqrt{y}|$.

Note that the Hamiltonian of (5.11) is indeed represented by this choice of objects. In particular, note that $\operatorname{grad}\mathcal{E}(x) = g^{-1}(x)\mathcal{E}'(x) = x(-\mu/x + 1) = x - \mu$.

Proof. The functional $\mathcal{E}$ is smooth and finite on $E_0 := (0, \infty)$. Working in the natural global chart, we can define a Riemannian metric using $g(x) = x^{-1}$, or equivalently $\langle v, w\rangle_{g(x)} := x^{-1}vw$ on the tangentbundle at x. This local metric indeed gives the global metric d of the lemma on $(0, \infty)$, which can then be extended by continuity to the boundary 0. We next verify the convexity of $\mathcal{E}$. As $\mathcal{E}(0) = \infty$, it suffices to consider geodesics that remain in $(0, \infty)$. Working infinitesimally and considering the geodesic from x to y, see Proposition 16.2 of [49], we verify

$$\begin{aligned}
&\langle -\operatorname{grad}\mathcal{E}(x), \operatorname{grad}_x \frac{1}{2}d^2(x, y)\rangle_{g(x)} - \langle -\operatorname{grad}\mathcal{E}(y), -\operatorname{grad}_y \frac{1}{2}d^2(x, y)\rangle_{g(y)} \\
&\quad = 2\,(\mu - x)\left(1 - \frac{\sqrt{y}}{\sqrt{x}}\right) - 2\left(-\,(\mu - y)\left(1 - \frac{\sqrt{y}}{\sqrt{x}}\right)\right) \\
&\quad = -2\left(\frac{\mu}{\sqrt{xy}} + 1\right)\left(\sqrt{x} - \sqrt{y}\right)^2 \\
&\leq -2\left(\sqrt{x} - \sqrt{y}\right)^2 = -\frac{1}{2}d^2(x, y),
\end{aligned}$$

implying that $\mathcal{E}$ is 1-convex. □

5.3. The Wasserstein space

We consider $E = \mathcal{P}_2(\mathbb{R}^d)$, which we equip with the Kantorovich-Wasserstein distance $W_2(\cdot,\cdot)$ of order two, defined by

$$W_2^2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int |x-y|^2 \pi(\mathrm{d}x\mathrm{d}y).$$

Following [3] we consider an energy functional $\mathcal{E}$ which is the sum of an internal energy, a potential energy and an interaction energy term. More precisely, we consider functions $F : \mathbb{R}_+ \to \mathbb{R}$, $V : \mathbb{R}^d \to \mathbb{R}$, $W : \mathbb{R}^d \to \mathbb{R}$ such that

Assumption 5.17 *(McCann's condition).*

(a) $F : [0,+\infty) \to \mathbb{R}$ is convex, differentiable with superlinear growth. It satisfies the doubling condition

$$\exists C > 0 : \quad F(z+w) \leq C(1 + F(z) + F(w)), \quad \forall z, w \geq 0.$$

Moreover we assume that

$$s \mapsto s^d F(s^{-d}) \quad \text{is convex and increasing on } (0,+\infty)$$

and

$$F(0) = 0, \quad \lim_{s\to 0} F(s)/s^{-\alpha} > -\infty, \quad \text{for some } \alpha > \frac{d}{d+2}.$$

(b) $V : \mathbb{R}^d \to (-\infty,+\infty]$ is lower semi-continuous, κ_V-convex for some $\kappa_V \in \mathbb{R}$, with proper domain that has nonempty interior.

(c) $W : \mathbb{R}^d \to [0,\infty)$ is an even continuously differentiable κ_W-convex function for some $\kappa_W \geq 0$[3] and satisfies the doubling condition

$$\exists C > 0 : \quad W(x+y) \leq C(1 + W(x) + W(y)), \quad \forall x, y \in \mathbb{R}^d.$$

We define our energy functional $\mathcal{E}$ by

$$\mathcal{E}(\rho) := \int F\Big(\frac{\mathrm{d}\rho}{\mathrm{d}\mathscr{L}^d}(x)\Big)\mathrm{d}x + \int V(x)\rho(\mathrm{d}x) + \frac{1}{2}\int W(x-y)\rho(\mathrm{d}x)\otimes\rho(\mathrm{d}y), \tag{5.12}$$

setting $\mathcal{E}(\rho) = +\infty$ as soon as ρ is not absolutely continuous w.r.t the Lebesgue measure $\mathscr{L}^d$. The gradient flow of functionals satisfying McCann's condition has attracted lots

[3] We impose $\kappa_W \geq 0$ as this condition allows us to directly apply the results of [3]. However, it is very likely that this assumption is not necessary and that $\kappa_W \in \mathbb{R}$ is enough for Theorem 5.18 to hold.

of interest over the past two decades, because of their connection with PDEs. Indeed, the gradient flow of Boltzmann's entropy $F(s) = s\log s$ provides with a variational interpretation of the heat equation [40], whereas the gradient flow of Rény's entropy ($F(s) = \frac{1}{\alpha-1}s^\alpha$) relates to the porous medium equation in the same way [44].

Theorem 5.18. *Let $(E,d) = (\mathcal{P}_2(\mathbb{R}^d), W_2(\cdot,\cdot))$ and $\mathcal{E}$ be defined by (5.12) with F, V, W satisfying Assumption 5.17. Then the conclusion of Theorem 2.14 hold with $\kappa = \kappa_V + \kappa_W$.*

The fact that the hypothesis of Theorem 2.14 are verified under Assumption 5.17 is a consequence of well-known results, that we essentially take from [3]. For the identification that $\kappa = \kappa_V + \kappa_W$, see Proposition 3.33 in [2].

Proof. We verify the hypothesis of Theorem 2.14 one by one.

- *Verification of 2.3* The completeness of $(\mathcal{P}_2(\mathbb{R}^d), W_2(\cdot,\cdot))$ is proven at [3, Prop. 7.1.5]. The fact that it is a geodesic space is proven at [2, Thm 2.10].
- *Verification of Assumption 2.5* The existence of an (EVI_κ) gradient flow on $\mathcal{P}_2(\mathbb{R}^d)$ is granted by [3, Theorems 11.2.1 and 11.2.8].
- *Verification of Assumption 2.9* Let us proceed to verify condition (2.6). Given ρ s.t. $I(\rho) < +\infty$ we know that against the Lebesgue measure ρ is regular in the sense of [3, Def. 6.2.2]. Thus, we can apply [3, Thm 6.2.4] to obtain the existence of a map $\mathbf{r}$ such that the (unique) geodesic $\boldsymbol{\zeta}^{\rho\to\pi}$ takes the form

$$\boldsymbol{\zeta}^{\rho\to\pi}(t) = (\boldsymbol{i} + t(\boldsymbol{r} - \boldsymbol{i}))_{\#}\rho \quad \forall t \in [0,1],$$

where $\boldsymbol{i}$ denotes the identity map. Moreover, thanks to [2, Thm 6.1 ii)] for any $\theta > 0$ we can find $\varphi^\theta \in C_c^\infty(\mathbb{R}^d)$ such that

$$|\nabla\varphi^\theta - (\mathbf{r} - \boldsymbol{i})|_{L^2_\rho} \leq \theta. \tag{5.13}$$

Using either a direct calculation or the isometry property of [2, Thm 6.1] we also find that if we define $\boldsymbol{\zeta}^{\rho\to\pi}_\theta(u) = (\boldsymbol{i} + u\nabla\varphi^\theta)_{\#}\rho$ for u small enough, then

$$\lim_{u\downarrow 0} \frac{W_2(\boldsymbol{\zeta}^{\rho\to\pi}_\theta(u), \boldsymbol{\zeta}^{\rho\to\pi}(u))}{u} \leq |\nabla\varphi^\theta - (\mathbf{r} - \boldsymbol{i})|_{L^2_\rho} \leq \theta,$$

which is (2.6). We now proceed to verify (2.7). By [3, Thm 10.4.13] we know that if $I(\rho) < +\infty$, then setting

$$L_F(z) = zF'(z) - F(z)$$

we have that $L_F\big(\frac{\mathrm{d}\rho}{\mathrm{d}\mathscr{L}^d}\big)$ belongs to $W^{1,1}_{loc}$. Combining [3, Lemma 10.4.4 and Eqs (10.4.58), (10.4.59)][4]

$$\begin{aligned}\lim_{u\downarrow 0}\frac{\mathcal{E}(\boldsymbol{\zeta}^{\rho\to\pi}_{\theta}(u))-\mathcal{E}(\rho)}{u} &= \int -L_F\big(\frac{\mathrm{d}\rho}{\mathrm{d}\mathscr{L}^d}\big)\Delta\varphi^{\theta}\mathrm{d}\mathscr{L}^d \\ &\quad + \int\langle\nabla V,\nabla\varphi^{\theta}\rangle\mathrm{d}\rho + \int\langle\nabla W*\rho,\nabla\varphi^{\theta}\rangle\mathrm{d}\rho \\ &= \int\Big\langle\frac{1}{\frac{\mathrm{d}\rho}{\mathrm{d}\mathscr{L}^d}}\nabla L_F\big(\frac{\mathrm{d}\rho}{\mathrm{d}\mathscr{L}^d}\big)+\nabla V+\nabla W*\rho,\nabla\varphi^{\theta}\Big\rangle\mathrm{d}\rho.\end{aligned}$$

Applying again [3, Thm 10.4.13] we have that there exist $\boldsymbol{w}\in L^2_{\rho}$ such that

$$\int|\boldsymbol{w}|^2\mathrm{d}\rho = I(\rho),$$
$$\boldsymbol{w} = \frac{1}{\frac{\mathrm{d}\rho}{\mathrm{d}\mathscr{L}^d}}\nabla L_F\big(\frac{\mathrm{d}\rho}{\mathrm{d}\mathscr{L}^d}\big)+\nabla V+\nabla W*\rho \quad \rho\text{-a.e.}$$

But then by Cauchy Schwartz we find

$$\lim_{u\downarrow 0}\frac{\mathcal{E}(\boldsymbol{\zeta}^{\rho\to\pi}_{\theta}(u))-\mathcal{E}(\rho)}{u}\leq\sqrt{I(\rho)}|\nabla\varphi^{\theta}|_{L^2_{\rho}}\leq\sqrt{I}(\rho)(W_2(\rho,\pi)+\theta),$$

where to obtain the last inequality we used (5.13), the triangular inequality and the fact that $W_2(\rho,\pi)=\int|\boldsymbol{r}-\boldsymbol{i}|^2\mathrm{d}\rho$. The proof of (2.7) is now complete. □

Data availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Appendix A

A.1. Ekeland's principle

Lemma A.1 *(Ekeland's principle). Let K be an abstract set and $\mathcal{B}:K\times K\to[0,+\infty)$ a function with the following properties:*

(i) $\mathcal{B}(x,x)=0$ *for all* $x\in K$
(ii) $\mathcal{B}(x,z)\leq\mathcal{B}(x,y)+\mathcal{B}(y,z)$ *for all* $x,y,z\in K$*.*

[4] In particular, one can check that the hypothesis of Lemma 10.4.4 are verified with $\boldsymbol{r}_t=(1-t)\boldsymbol{i}+t\nabla\varphi^{\theta}$ using, among other things, the fact that for t small enough $\boldsymbol{r}_t$ is invertible, smooth, strongly convex and $(\boldsymbol{r}_t)_{\#}\rho\ll\mathscr{L}^d$.

(iii) For any sequence $(x_n)_{n\in\mathbb{N}} \in K$ *satisfying* $\sum_{n\in\mathbb{N}} \mathcal{B}(x_{n+1}, x_n) < +\infty$*, there exists* $x \in K$ *such that* $\lim_{n\to\infty} \mathcal{B}(x, x_n) = 0$*.*

Let $\mathcal{G} : K \to [-\infty, +\infty)$ *be a bounded from above function, i.e.* $\sup_{x\in K} \mathcal{G}(x) < +\infty$*, such that:*

- *if* $(x_n)_{n\in\mathbb{N}}, x \in K$*,* $\sum_{n\in\mathbb{N}} \mathcal{B}(x_{n+1}, x_n) < +\infty$ *and* $\lim_{n\to+\infty} \mathcal{B}(x, x_n) = 0$ *then*

$$\mathcal{G}(x) \geq \limsup_n \mathcal{G}(x_n).$$

Then for each $\delta > 0$ *and any* $\hat{x} \in K$ *such that* $\mathcal{G}(\hat{x}) \neq -\infty$ *there exists* $x_\delta \in K$ *such that*

(1) $\mathcal{G}(\hat{x}) + \frac{1}{2}\delta\mathcal{B}(x_\delta, \hat{x}) \leq \mathcal{G}(x_\delta)$*,*
(2) $\sup_x \left\{\mathcal{G}(x) - \frac{1}{2}\delta\mathcal{B}(x, x_\delta)\right\} \leq \mathcal{G}(x_\delta)$*.*

Let us note as a corollary that the above statements have the following consequences

(a) Suppose that $\mathcal{G}(\hat{x}) \geq \sup_{x\in K} \mathcal{G}(x) - \frac{1}{2}\delta^2$*, then* $\mathcal{B}(x_\delta, \hat{x}) \leq \delta$*.*
(b) For all $x \neq x_\delta$ *we have* $\mathcal{G}(x) - \delta\mathcal{B}(x, x_\delta) < \mathcal{G}(x_\delta)$*.*
(c) Suppose that $(x_n)_{n\in\mathbb{N}} \in K$ *is such that* $\lim_{n\to\infty} \mathcal{G}(x_n) - \delta\mathcal{B}(x_n, x_\delta) = \mathcal{G}(x_\delta)$*, then*

$$\lim_{n\to\infty} \mathcal{B}(x_n, x_\delta) = 0 \quad \text{and} \quad \lim_{n\to\infty} \mathcal{G}(x_n) = \mathcal{G}(x_\delta).$$

Remark A.2. In particular, from (1) we deduce $\mathcal{G}(x_\delta) > -\infty$ and from (b) we deduce that x_δ is the unique optimizer of $\mathcal{G}(x) - \delta\mathcal{B}(x, x_\delta)$.

Proof. The statements (1) and (2) follow as in [47], using as $\mathcal{B}(x, y) := \mathcal{B}(y, x)$, $u(x) := -\mathcal{G}(x)$, multiplying all terms by -1 and replacing δ by $\frac{1}{2}\delta$. From (1) and (2), the consequences (a) and (b) follow immediately. We are left to prove (c).

Let $(x_n)_{n\in\mathbb{N}} \in K$ be as in (c). Then by statement (2), we have

$$0 \leq \mathcal{G}(x_\delta) - \mathcal{G}(x_n) + \frac{1}{2}\delta\mathcal{B}(x_n, x_\delta).$$

Thus,

$$\begin{aligned} 0 &\leq \frac{1}{2}\delta\mathcal{B}(x_n, x_\delta) \\ &\leq \mathcal{G}(x_\delta) - \mathcal{G}(x_n) + \delta\mathcal{B}(x_n, x_\delta). \end{aligned}$$

By assumption, the right hand side converges to 0. Therefore, we also have

$$\lim_{n\to\infty} \mathcal{B}(x_n, x_\delta) = 0.$$

Using again (2),

$$\mathcal{G}(x_\delta) \geq \mathcal{G}(x_n) - \frac{1}{2}\delta\mathcal{B}(x_n, x_\delta) \geq \limsup_{n\to\infty} \mathcal{G}(x_n) - \frac{1}{2}\delta\mathcal{B}(x_n, x_\delta) = \limsup_{n\to\infty} \mathcal{G}(x_n).$$

Moreover, by the assumption on the sequence $(x_n)_{n\in\mathbb{N}}$, we also have

$$\liminf_{n\to\infty} \mathcal{G}(x_n) \geq \liminf_{n\to\infty} \mathcal{G}(x_n) - \delta\mathcal{B}(x_n, x_\delta) = \mathcal{G}(x_\delta).$$

We then conclude that $\lim_{n\to\infty} \mathcal{G}(x_n) = \mathcal{G}(x_\delta)$. □

Let us show in the following lemma that Ekeland's principle can be applied to the Tataru distance.

Lemma A.3. *The Tataru distance satisfies the assumptions of Lemma A.1.*

Proof. (i) is trivial and (ii) has been verified in Lemma 4.4. Let us show (iii).

Let $(\mu_n)_{n\in\mathbb{N}} \in E$ be such that $\sum_n d_T(\mu_{n+1}, \mu_n) < \infty$.

Recall we have seen that

$$d_T(\mu, \nu) = \min_{t\geq 0} \left\{t + e^{\hat{\kappa}t} d(\mu, \nu(t))\right\}$$

Thus, there exists a sequence $(t_n)_{n\in\mathbb{N}} \in [0, +\infty)$ such that

$$\sum_n t_n + e^{\hat{\kappa}t_n} d(\mu_{n+1}, \mu_n(t_n)) < \infty. \tag{A.1}$$

For all $n \in \mathbb{N}$, set $s_n := \sum_{k=n}^{\infty} t_k$. Note that (A.1) implies that

$$s_n \leq s_0 = \sum_n t_n =: T < \infty,$$

$$\sum_n d(\mu_{n+1}, \mu_n(t_n)) = \sum_n e^{-\hat{\kappa}t_n} e^{\hat{\kappa}t_n} d(\mu_{n+1}, \mu_n(t_n))$$
$$\leq e^{-\hat{\kappa}T} \sum_n e^{\hat{\kappa}t_n} d(\mu_{n+1}, \mu_n(t_n)) < \infty. \tag{A.2}$$

(Remember that $\hat{\kappa} \leq 0$.)

Let us consider the sequence $(\nu_n)_{n\in\mathbb{N}} \in [0, +\infty) \in E$ given by $\nu_n := \mu_n(s_n)$ for all $n \in \mathbb{N}$. It follows by equation (4.3) that

$$\begin{aligned}\sum_n d(\nu_n,\nu_{n+1}) &= \sum_n d(\mu_n(s_n),\mu_{n+1}(s_{n+1}))\\ &\le \sum_n e^{-\kappa s_{n+1}} d(\mu_n(t_n),\mu_{n+1})\\ &\le \sum_n e^{-\hat\kappa s_{n+1}} d(\mu_n(t_n),\mu_{n+1})\\ &\le e^{-\hat\kappa T}\sum_n d(\mu_n(t_n),\mu_{n+1})\\ &< \infty.\end{aligned}$$

Therefore $(\nu_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and converges to a $\nu \in E$, i.e.

$$\lim_{n\to\infty} d(\nu,\mu_n(s_n)) = 0.$$

Moreover

$$\begin{aligned}0 \le \lim_{n\to\infty} d_T(\nu,\mu_n) &= \lim_{n\to\infty}\inf_{t\ge 0}\left\{t + e^{\hat\kappa t} d(\nu,\mu_n(t))\right\}\\ &\le \lim_{n\to\infty} s_n + e^{\hat\kappa s_n} d(\nu,\mu_n(s_n)) = 0. \quad \square\end{aligned}$$

A.2. From optimizing sequences to optimizing points

The following Lemma relates Definition 2.13 to the classical definition stated in terms of optimizing points. We use the lemma in combination with Ekeland's principle in the proof of the comparison principle.

Lemma A.4. *Consider a viscosity subsolution u of equation (2.10). Let $(f,g) \in A_\dagger$ and $(\pi_n)_{n\in\mathbb{N}} \in E$, be the sequence given by the definition of viscosity subsolution. Suppose that:*

- *There exists $\pi_0 \in E$ such that $\lim_n \pi_n = \pi_0$ and*

$$u(\pi_0) - f(\pi_0) = \sup_\pi u(\pi) - f(\pi).$$

Then we have

$$u(\pi_0) - \lambda g(\pi_0) - h^\dagger(\pi_0) \le 0.$$

Consider a viscosity supersolution v of equation (2.11). Let $(f,g) \in A_\ddagger$ and $(\pi_n)_{n\in\mathbb{N}} \in E$, be the sequence given by the definition of viscosity supersolution. Suppose that:

- *There exists* $\pi_0 \in E$ *such that* $\lim_n \pi_n = \pi_0$ *and*

$$v(\pi_0) - f(\pi_0) = \inf_{\pi} v(\pi) - f(\pi).$$

Then we have

$$v(\pi_0) - \lambda g(\pi_0) - h^{\ddagger}(\pi_0) \geq 0.$$

Proof. We prove the statement for the subsolution case, the supersolution case works analogously.

Let u be a subsolution to $f - \lambda A_{\dagger} f = h^{\dagger}$, $(f, g) \in A_{\dagger}$ and $(\pi_n)_{n\in\mathbb{N}} \in E$ be as in the assumption of this lemma. Then in particular we have

$$\limsup_n u(\pi_n) - f(\pi_n) = \sup_{\pi} u(\pi) - f(\pi),$$
$$\limsup_n u(\pi_n) - \lambda g(\pi_n) - h^{\dagger}(\pi_n) \leq 0.$$

By assumption, there exists $\pi_0 \in E$ such that $u(\pi_0) - f(\pi_0) = \sup_{\pi} u(\pi) - f(\pi)$ and $\pi_n \to \pi_0$.

Being u upper semi-continuous, we have $\limsup_n u(\pi_n) \leq u(\pi_0)$. On the other hand, being $\lim_n u(\pi_n) - f(\pi_n) = u(\pi_0) - f(\pi_0)$, we have

$$\begin{aligned}\liminf_n u(\pi_n) &= \liminf_n (u(\pi_n) - f(\pi_n) + f(\pi_n)) \\ &\geq u(\pi_0) - f(\pi_0) + \liminf_n f(\pi_n) \\ &\geq u(\pi_0) - f(\pi_0) + f(\pi_0) = u(\pi_0)\end{aligned}$$

due to the fact that f is continuous. We can then conclude that $\lim_n u(\pi_n) = u(\pi_0)$. On the other hand, being $h^{\dagger}$ continuous and g is upper semi-continuous, we find

$$\begin{aligned}0 &\geq \limsup_n (u(\pi_n) - \lambda g(\pi_n) - h^{\dagger}(\pi_n)) \\ &= u(\pi_0) - h^{\dagger}(\pi_0) + \limsup_n -\lambda g(\pi_n) \\ &\geq u(\pi_0) - h^{\dagger}(\pi_0) + \liminf_n -\lambda g(\pi_n) \\ &= u(\pi_0) - h^{\dagger}(\pi_0) - \lambda \limsup_n g(\pi_n) \\ &\geq u(\pi_0) - h^{\dagger}(\pi_0) - \lambda g(\pi_0). \quad \square\end{aligned}$$

A.3. A variant of the triangle inequality for the quadratic distance

For the proof of Proposition 3.2, we need the following combination of the triangle and Jensen inequality.

Lemma A.5. *Let $\nu_1, \nu_2, \nu_3, \nu_4 \in E$ and $\varepsilon, \varepsilon' \in (0, 1/3)$, then*

$$\frac{1}{6}\frac{1}{1-\varepsilon'}\frac{1}{2}d^2(\nu_1,\nu_4) \leq \frac{1}{1-\varepsilon}\frac{1}{2}d^2(\nu_1,\nu_2) + \frac{1}{2}d^2(\nu_2,\nu_3) + \frac{1}{1+\varepsilon}\frac{1}{2}d^2(\nu_3,\nu_4) \tag{A.3}$$

Proof. By the triangle inequality, we have

$$d(\nu_1,\nu_4) \leq d(\nu_1,\nu_2) + d(\nu_2,\nu_3) + d(\nu_3,\nu_4)$$

so that by Jensens inequality, we have

$$\begin{aligned}\frac{1}{6}d^2(\nu_1,\nu_4) &\leq \frac{1}{3}\frac{1}{2}\left(d(\nu_1,\nu_2) + d(\nu_2,\nu_3) + d(\nu_3,\nu_4)\right)^2 \\ &= 3\frac{1}{2}\left(\frac{1}{3}d(\nu_1,\nu_2) + \frac{1}{3}d(\nu_2,\nu_3) + \frac{1}{3}d(\nu_3,\nu_4)\right)^2 \\ &\leq \frac{3}{2}\left(\frac{1}{3}d^2(\nu_1,\nu_2) + \frac{1}{3}d^2(\nu_2,\nu_3) + \frac{1}{3}d^2(\nu_3,\nu_4)\right).\end{aligned}$$

The second claim follows from this inequality, using that for $\varepsilon, \varepsilon' \in (0, 1/3)$

$$\frac{1-\varepsilon}{1-\varepsilon'} \leq 2, \qquad \frac{1}{2(1-\varepsilon')} \leq 1, \qquad 1 \leq \frac{1}{1-\varepsilon}, \qquad \frac{1+\varepsilon}{1-\varepsilon'} \leq 2. \qquad \square$$

References

[1] L. Ambrosio, J. Feng, On a class of first order Hamilton-Jacobi equations in metric spaces, J. Differ. Equ. 256 (7) (2014) 2194–2245.

[2] L. Ambrosio, N. Gigli, A user's guide to optimal transport, in: Modelling and Optimisation of Flows on Networks, Springer, 2013, pp. 1–155.

[3] L. Ambrosio, N. Gigli, G. Savaré, Gradient Flows: in Metric Spaces and in the Space of Probability Measures, Springer Science & Business Media, 2008.

[4] E. Bandini, et al., Randomized filtering and Bellman equation in Wasserstein space for partial observation control problem, Stoch. Process. Appl. 129 (2) (2019) 674–711.

[5] L. Bertini, et al., Macroscopic fluctuation theory for stationary non-equilibrium states, J. Stat. Phys. 107 (3-4) (2002) 635–675.

[6] L. Bertini, et al., Macroscopic fluctuation theory, Rev. Mod. Phys. 87 (2) (2015) 593–636.

[7] A. Bensoussan, J. Frehse, P. Yam, et al., Mean Field Games and Mean Field Type Control Theory, vol. 101, Springer, 2013.

[8] A. Bensoussan, P.J. Graber, S.C.P. Yam, Control on Hilbert spaces and application to mean field type control theory, preprint, arXiv:2005.10770, 2020.

[9] M. Burzoni, et al., Viscosity solutions for controlled McKean–Vlasov jump-diffusions, SIAM J. Control Optim. 58 (3) (2020) 1676–1699.

[10] P. Cannarsa, C. Sinestrari, Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control, vol. 58, Springer Science & Business Media, 2004.

[11] P. Cardaliaguet, et al., The Master Equation and the Convergence Problem in Mean Field Games: (AMS-201), vol. 201, Princeton University Press, 2019.

[12] P. Cardaliaguet, M. Cirant, A. Porretta, Splitting methods and short time existence for the master equations in mean field games, J. Eur. Math. Soc. Online first (2022).

[13] R. Carmona, F. Delarue, et al., Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics, Ann. Probab. 43 (5) (2015) 2647–2700.

[14] R. Carmona, F. Delarue, A. Lachapelle, Control of McKean–Vlasov dynamics versus mean field games, Math. Financ. Econ. 7 (2) (2013) 131–166.

[15] M.G. Crandall, H. Ishii, P.-L. Lions, User's guide to viscosity solutions of second order partial differential equations, Bull. Am. Math. Soc. (N.S.) 27 (1) (1992) 1–67.
[16] M.G. Crandall, P.-L. Lions, Hamilton-Jacobi Equations in Infinite Dimensions. Part 1. Uniqueness of Viscosity Solutions, Tech. Rep., Wisconsin Univ-Madison Mathematics Research Center, 1984.
[17] M.G. Crandall, P.-L. Lions, Hamilton-Jacobi equations in infinite dimensions. II. Existence of viscosity solutions, J. Funct. Anal. 65 (3) (1986) 368–405.
[18] M.G. Crandall, P.-L. Lions, Hamilton-Jacobi equations in infinite dimensions, III, J. Funct. Anal. 68 (2) (1986) 214–247.
[19] M.G. Crandall, P.-L. Lions, Viscosity solutions of Hamilton-Jacobi equations in infinite dimensions. IV. Hamiltonians with unbounded linear terms, J. Funct. Anal. 90 (2) (1990) 237–283.
[20] M.G. Crandall, P.-L. Lions, Viscosity solutions of Hamilton-Jacobi equations in infinite dimensions. V. Unbounded linear terms and B-continuous solutions, J. Funct. Anal. 97 (2) (1991) 417–465.
[21] M.G. Crandall, P.-L. Lions, Hamilton-Jacobi equations in infinite dimensions. VI. Nonlinear A and Tataru's method refined, in: Evolution Equations, Control Theory, and Biomathematics, Han sur Lesse, 1991, in: Lecture Notes in Pure and Appl. Math., vol. 155, Dekker, New York, 1994, pp. 51–89.
[22] A. Cosso, et al., Master Bellman equation in the Wasserstein space: uniqueness of viscosity solutions, preprint, arXiv:2107.10535, 2021.
[23] J.C. Cox, J.E. Ingersoll, S.A. Ross, A theory of the term structure of interest rates, Econometrica 53 (2) (1985) 385–407.
[24] S. Daneri, G. Savaré, Eulerian calculus for the displacement convexity in the Wasserstein distance, SIAM J. Math. Anal. 40 (3) (2008) 1104–1122.
[25] X. Deng, J. Feng, Y. Liu, A singular 1-D Hamilton-Jacobi equation, with application to large deviation of diffusions, Commun. Math. Sci. 9 (2011) 1.
[26] J. Dolbeault, B. Nazaret, G. Savaré, A new class of transport distances between measures, Calc. Var. Partial Differ. Equ. 34 (2) (2009) 193–231.
[27] J. Feng, Large deviation for diffusions and Hamilton-Jacobi equation in Hilbert spaces, Ann. Probab. 34 (1) (2006) 321–385.
[28] J. Feng, M. Katsoulakis, A comparison principle for Hamilton-Jacobi equations related to controlled gradient flows in infinite dimensions, Arch. Ration. Mech. Anal. 192 (2) (2009) 275–310.
[29] J. Feng, T.G. Kurtz, Large Deviations for Stochastic Processes, American Mathematical Society, 2006, xii+410.
[30] J. Feng, T. Mikami, J. Zimmer, A Hamilton-Jacobi PDE associated with hydrodynamic fluctuations from a nonlinear diffusion equation, Commun. Math. Phys. 385 (1) (2021) 1–54.
[31] J. Feng, T. Nguyen, Hamilton–Jacobi equations in space of measures associated with a system of conservation laws, J. Math. Pures Appl. 97 (4) (2012) 318–390.
[32] P. Fuchs, A. Jüngel, M.-K. von Renesse, On the Lagrangian structure of quantum fluid models, Discrete Contin. Dyn. Syst. 34 (4) (2014) 1375–1396.
[33] W. Gangbo, et al., Mean field games master equations with non-separable Hamiltonians and displacement monotonicity, Ann. Probab. 50 (6) (2022) 2178–2217.
[34] W. Gangbo, A.R. Mészáros, Global well-posedness of Master equations for deterministic displacement convex potential mean field games, Comm. Pure Appl. Math 75 (12) (2022) 2685–2801.
[35] W. Gangbo, A. Święch, Existence of a solution to an equation arising from the theory of mean field games, J. Differ. Equ. 259 (11) (2015) 6573–6643.
[36] W. Gangbo, A. Święch, Metric viscosity solutions of Hamilton–Jacobi equations depending on local slopes, Calc. Var. Partial Differ. Equ. 54 (1) (2015) 1183–1218.
[37] W. Gangbo, A. Tudorascu, On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations, J. Math. Pures Appl. 125 (2019) 119–174.
[38] I. Gentil, C. Léonard, L. Ripani, Dynamical aspects of the generalized Schrödinger problem via Otto calculus–a heuristic point of view, Rev. Mat. Iberoam. 36 (4) (2020) 1071–1112.
[39] N. Gozlan, C. Roberto, P.-M. Samson, Hamilton Jacobi equations on metric spaces and transport entropy inequalities, Rev. Mat. Iberoam. 30 (1) (2014) 133–163.
[40] R. Jordan, D. Kinderlehrer, F. Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1) (1998) 1–17.
[41] R.C. Kraaij, F. Redig, R. Versendaal, Classical large deviation theorems on complete Riemannian manifolds, Stoch. Process. Appl. 129 (11) (2019) 4294–4334.
[42] L. Monsaingeon, L. Tamanini, D. Vorotnikov, The dynamical Schrödinger problem in abstract metric spaces, preprint, arXiv:2012.12005, 2020.
[43] M. Muratori, G. Savaré, Gradient flows and evolution variational inequalities in metric spaces. I: structural properties, J. Funct. Anal. 278 (4) (2020) 108347.

[44] F. Otto, The Geometry of Dissipative Evolution Equations: the Porous Medium Equation, 2001.
[45] H. Pham, X. Wei, Bellman equation and viscosity solutions for mean-field stochastic control problem, ESAIM Control Optim. Calc. Var. 24 (1) (2018) 437–461.
[46] M.-K. von Renesse, An optimal transport view of Schrödinger's equation, Can. Math. Bull. 55 (4) (2012) 858–869.
[47] D. Tataru, Viscosity solutions of Hamilton-Jacobi equations with unbounded nonlinear terms, J. Math. Anal. Appl. 163 (2) (1992) 345–392.
[48] D. Tataru, Viscosity solutions for Hamilton-Jacobi equations with unbounded nonlinear term: a simplified approach, J. Differ. Equ. 111 (1) (1994) 123–146.
[49] C. Villani, Optimal Transport, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences. Old and new), vol. 338, Springer-Verlag, Berlin, 2009, xxii+973.
[50] C. Wu, J. Zhang, Viscosity solutions to parabolic master equations and McKean–Vlasov SDEs with closed-loop controls, Ann. Appl. Probab. 30 (2) (2020) 936–986.