Hamilton-Jacobi equations on infinite dimensional spaces corresponding to linearly controlled gradient flows of an energy functional

Daniela Tonon
a collaboration with Giovanni Conforti (École Polytechnique) and Richard Kraaij (TU Delft)

Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova

GdR MOA, Nice
October 13th 2022

DIPARTIMENTO MATEMATICA
DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

800 ANNI 1222·2022
UNIVERSITÀ DEGLI STUDI DI PADOVA

The Hamilton-Jacobi equation

Given a metric space (E, d), where the metric d is generated by a Riemannian metric $\langle\cdot,\cdot\rangle$, we consider a stationary Hamilton-Jacobi (HJ) equation of the following type

$$f - \lambda Hf = h,$$

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

where *grad* is the gradient associated with the metric and $\mathcal{E}$ an energy functional

Difficulties: the energy functional $\mathcal{E}$ does not have compact sublevel sets and a precise notion of its gradient is usually difficult or impossible to give

AIM: prove a comparison principle for viscosity solutions of the above HJ equation

A fundamental example

- $E = \mathcal{P}_2(\mathbb{R}^d)$ is the space of probability measures with finite second moment equipped with the Riemannian metric of optimal transport (Otto metric) that generates the Wasserstein distance of order two W_2,
- $\mathcal{E}(\mu) = \int \mu \log \mu$ an entropic functional satisfying *EVI* condition, possibly including an interaction term

The HJ equation is expected to characterize the value function

$$f(\rho_0) = \sup \left\{ \int_0^{+\infty} e^{-\lambda^{-1} t} \left[\lambda^{-1} h(\rho^u(t)) - \frac{1}{2} \|u(t)\|^2 \right] \mathrm{d}t \right\},$$

over all absolutely continuous curves $(\rho^u(s))_{s \in [0,+\infty]} \subset \mathcal{P}_2(\mathbb{R}^d)$ that are weak solutions of the following continuity equation

$$\dot{\rho}^u = -\mathrm{grad}^{W_2} \mathcal{E}(\rho^u) + u, \quad \rho^u(0) = \rho_0$$

for a control $u(s) \in T_{\rho^u(s)} \mathcal{P}_2(\mathbb{R}^d)$

A fundamental example

- $E = \mathcal{P}_2(\mathbb{R}^d)$ is the space of probability measures with finite second moment equipped with the Riemannian metric of optimal transport (Otto metric) that generates the Wasserstein distance of order two W_2,
- $\mathcal{E}(\mu) = \int \mu \log \mu$ an entropic functional satisfying *EVI* condition, possibly including an interaction term

The HJ equation is expected to characterize the value function

$$f(\rho_0) = \sup \left\{ \int_0^{+\infty} e^{-\lambda^{-1} t} \left[\lambda^{-1} h(\rho^u(t)) - \frac{1}{2} \|u(t)\|^2 \right] \mathrm{d}t \right\},$$

over all absolutely continuous curves $(\rho^u(s))_{s\in[0,+\infty]} \subset \mathcal{P}_2(\mathbb{R}^d)$ that are weak solutions of the following continuity equation

$$\partial_t \rho^u - \frac{1}{2}\Delta\rho^u + \nabla\cdot(\rho^u u) = 0, \quad \rho^u(0) = \rho_0$$

for a control $u(s) \in T_{\rho^u(s)}\mathcal{P}_2(\mathbb{R}^d)$

Hamilton-Jacobi equations on infinite dimensional spaces

The starting point: the articles of Crandall and Lions '84 on infinite dimensional HJ equations in Hilbert spaces or Banach spaces with Randon-Nykodim property

Our setting: HJ equations on metric spaces that are not necessarily Hilbert, and in particular over the space of probability measures $\mathcal{P}_2(\mathbb{R}^d)$ endowed with a transport-like distance

Motivation: The mean field Schrödinger problem and recent applications in large deviations, functional inequalities, statistical mechanics, and McKean-Vlasov control

Different strategies in the case $E = \mathcal{P}_2(\mathbb{R}^d)$:

- Lifting of functions:
  The idea consists in associating to any $v : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ a function V defined on $L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$ by setting for any random variable $X \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$

  $$V(X) = v(\mu)$$

  where $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ is such that $\mu = Law(X)$

  As a derivative one can use Lions derivative exploiting the Hilbert space properties of $L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$

  (Bandini et al. 2019, Bensoussan, Graber and Yam 2020, Pham and Wei 2018, etc.)

Several different strategies:

- Intrinsic approach:
  It consists in working directly at the level of the space of probability measures and develop all the relevant notions therein
  - use the linear derivative, as shown in the context of McKean-Vlasov control for jump processes (Burzoni et al. 2020)
  - use the notion of derivative on Wasserstein spaces that comes form optimal transport theory (Ambrosio Gigli Savaré 2008)

The connections between the intrinsic approach and the extrinsic notion of derivative obtained through the lifting procedure have been clarified in Gangbo and Tudorascu 2019

Intrinsic approach: other important contributions (Feng and collaborators since 2006, Gangbo and collaborators since 2015, Wu and Zhang 2020)

Crucial assumption: the variations of the Hamiltonian w.r.t. the measure argument can be locally controlled by the metric in some way

Again on the fundamental example:

We recall: $\mathcal{E}$ is the relative entropy and (E, d) the Wasserstein space

$$f - \lambda H f = h,$$

with $Hf(\pi) := -\langle \mathrm{grad}^{W_2}\mathcal{E}, \mathrm{grad}^{W_2} f\rangle(\pi) + \frac{1}{2}\|\mathrm{grad}^{W_2} f\|^2(\pi)$

Setting $\lambda = 1$, the formal change of variable $\tilde{f} = f - \mathcal{E}$ allows to rewrite our equation as

$$f(\pi) - \frac{1}{2}\|\mathrm{grad}^{W_2} f\|^2(\pi) + \mathcal{F}(\pi) = 0,$$

where $\mathcal{F}(\pi) := \frac{1}{2}\|\mathrm{grad}^{W_2}\mathcal{E}\|^2(\pi) + \mathcal{E}(\pi) - h(\pi)$

PROBLEM: the assumption of the existence of an EVI gradient flow, formally equivalent to the semiconvexity of $\mathcal{E}$, does not give any control on the growth of $\|\mathrm{grad}^{W_2}\mathcal{E}\|^2(\pi)$

$\Longrightarrow$ the techniques developed in the mentioned references cannot be successfully applied to our case

The abstract setting

We consider the following stationary HJ equation on (E, d)

$$f - \lambda Hf = h$$

where $\lambda > 0$ is a constant, h a real function in $C_b(E)$ and

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

where

- (E, d) has to be geodesic space

  $\implies \forall \rho, \pi \in E$ there exists a geodesic, i.e. $(\gamma_\rho^\pi(t))_{t\in[0,1]}$ s.t. $\gamma_\rho^\pi(0) = \rho, \gamma_\rho^\pi(1) = \pi$ and $\forall s, t \in [0, 1]$

  $$d(\gamma_\rho^\pi(s), \gamma_\rho^\pi(t)) = |t - s| d(\rho, \pi)$$

- $\mathcal{E} : E \to (-\infty, +\infty]$ is an extended, l.s.c., energy functional with proper effective domain, dense in E

  $$\mathcal{D}(\mathcal{E}) := \{\pi \in E : \mathcal{E}(\pi) < +\infty\} \neq \emptyset, \quad \overline{\mathcal{D}(\mathcal{E})} = E$$

NOTE: $\mathcal{E}$ and d do not necessarily have compact sublevel sets. In fact, we allow $\mathcal{E}$ that are unbounded from below

NOTE: In the typical situations of interest, when $(E,d) = (\mathcal{P}_2(\mathbb{R}^d), W_2)$, $\mathcal{E}$ is $+\infty$ on a dense set and nowhere differentiable, even though the subdifferential is well defined and non empty on a subset of the domain of $\mathcal{E}$

WARNING: a precise notion of gradient for $\mathcal{E}$ is usually difficult or impossible to give

We use the notion of local slope

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

formally $|\partial\phi(\rho)|$ can be seen as the norm of the gradient of ϕ

Hypotheses on $\mathcal{E}$:

Angle condition: any geodesic $\gamma^\pi_\rho(\cdot)$ can be approximated for any $\theta > 0$ with a smoother curve $(\gamma^\pi_\rho)_\theta(\cdot)$,

$$\limsup_{t\downarrow 0} \frac{d((\gamma^\pi_\rho)_\theta(t), \gamma^\pi_\rho(t))}{t} \leq \theta,$$

along which the variations of $\mathcal{E}$ can be controlled with its local slope

$$\liminf_{t\downarrow 0} \frac{\mathcal{E}((\gamma^\pi_\rho)_\theta(t)) - \mathcal{E}(\rho)}{t} \leq |\partial\mathcal{E}|(\rho)(d(\rho,\pi) + \theta)$$

Remark: The angle condition is coherent with the interpretation of the local slope as the norm of the gradient of $\mathcal{E}$ and can be interpreted as the controllability of directional derivatives of the energy functional along regularized geodesics by its local slope

Hypotheses on $\mathcal{E}$:

Existence of an EVI_κ gradient flow defined on E for a given $\kappa \in \mathbb{R}$:

There exists a family of continuous maps $S(t) : E \to E$, $E = \overline{\mathcal{D}(\mathcal{E})}$, such that for all $\pi \in E$, the semigroup property holds

$$S[\pi](0) = \pi, \quad S[\pi](t+s) = S[S[\pi](t)](s) \ \forall t, s \in (0, +\infty)$$

and the curve $(S[\pi](t))_{t \geq 0}$ satisfies the following EVI_κ inequality:

$$\forall \rho \in \mathcal{D}(\mathcal{E}), \forall t \in [0, +\infty)$$

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\left(d^2(S[\pi](t), \rho)\right) \leq \mathcal{E}(\rho) - \mathcal{E}(S[\pi](t)) - \frac{\kappa}{2} d^2(S[\pi](t), \rho)$$

Remarks

- We allow also a possibly negative contractivity constant κ in *EVI*, i.e. a negatively curved space
- *EVI* implies uniqueness of the gradient flow
- *EVI* gradient flows enjoy useful regularizing properties that include energy dissipation and distance contraction estimates
- $S[\pi](t)$ solves formally $\frac{\mathrm{d}}{\mathrm{d}t}S[\pi](t) = -grad\ \mathcal{E}(S[\pi](t))$
- the Hamiltonian may be written as

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

Some properties:

- For all $\pi, \rho \in E$ the metric d formally satisfies

$$\left|\partial_\pi \left(\frac{1}{2} d^2(\pi,\rho)\right)\right|^2 = d^2(\pi,\rho)$$

- Let $\mu, \nu \in E$ and let $(S[\mu](t))_{t\geq 0}$,$(S[\nu](t))_{t\geq 0}$ be the corresponding gradient flow. Then we have

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

Some properties:

- For each $c_1 > -\kappa$ and for each $\nu \in E$ there exist $c_2, \tilde{c}_2 \in \mathbb{R}$ s.t.

$$\mathcal{E}(\pi) \geq -\frac{c_1}{2} d^2(\pi, \nu) - c_2,$$

i.e. 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

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

Moreover

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

EVI and displacement κ-convexity

Definition: Let $\kappa \in \mathbb{R}$. We say that a lower semi-continuous functional $\mathcal{E} : E \to \mathbb{R} \cup +\infty$ is strongly displacement κ-convex if for all geodesic $(\gamma(t))_{0 \le t \le 1} \subset E$ we have

$$\mathcal{E}(\gamma(t)) \le (1-t)\mathcal{E}(\gamma(0)) + t\mathcal{E}(\gamma(1)) - \frac{\kappa}{2} t(1-t) d^2(\gamma(0), \gamma(1)) \text{ for all } t \in [0,1]$$

Theorem: Daneri-Savaré 2008 Consider a lower semi-continuous functional $\mathcal{E} : E \to \mathbb{R} \cup +\infty$ on a geodesic space (E, d) such that there exists a gradient flow satisfying (EVI_κ) then $\mathcal{E}$ is strongly displacement κ-convex

REM the converse is not proved in general but it is true in most of the relevant examples (Hilbert spaces, Wasserstein spaces,...)

Our approach for the comparison theorem

Our strategy comes from a powerful approach initiated by Crandall and Lions, Tataru, Feng and collaborators that exploits the geometry of the underlying control problem and the *EVI* inequality

MAIN IDEAS:

- instead of working directly with H, we construct, via *EVI* inequality, 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 deal with the non compactness of the space we use of Ekeland's optimization principle
- to deal with the arising drift term we use Tataru's distance as a penalization function

Definition of viscosity super and sub solutions

Given the stationary HJ equation

$$f - \lambda H f = h$$

Definition: We say that $u : E \to \mathbb{R}$ is a (viscosity) subsolution if u is bounded, upper semi-continuous and if for all $f_1 \in \mathcal{D}(H)$ and $\rho_0 \in E$ s.t.

$$u(\rho_0) - f_1(\rho_0) = \sup_{\rho} u(\rho) - f_1(\rho)$$

we have

$$u(\rho_0) - \lambda H f_1(\rho_0) - h(\rho_0) \leq 0$$

Definition of viscosity super and sub solutions

Given the stationary HJ equation

$$f - \lambda Hf = h$$

Definition: We say that $v : E \to \mathbb{R}$ is a (viscosity) supersolution if v is bounded, lower semi-continuous and if for all $f_2 \in \mathcal{D}(H)$ and $\rho_0 \in E$ s.t.

$$v(\rho_0) - f_2(\rho_0) = \inf_{\rho} v(\rho) - f_2(\rho),$$

we have

$$v(\rho_0) - \lambda Hf_2(\rho_0) - h(\rho_0) \geq 0$$

The comparison principle

We aim at a comparison principle of this type:

Let u be a subsolution to $f - \lambda Hf = h_1$ and let v be a supersolution to $f - \lambda Hf = h_2$. Then we have

$$\sup_{\mu} u(\mu) - v(\mu) \leq \sup_{\mu} h_1(\mu) - h_2(\mu)$$

REM: In general, the comparison principle proof relies upon test functions which behave like the square of distance functions

For instance, in the $\mathbb{R}^d$ case, these test functions take the form $\frac{1}{2}|x-y|^2$

NOTE: In the infinite dimensional case, functions like d^2 are not necessarily included in the domain of the Hamiltonian

IDEA: if $f_1(\pi) = \frac{1}{2} a d^2(\pi, \rho)$ for some $\rho \in E$ and $a > 0$, then formally

$$Hf_1(\pi) = \frac{1}{2} a \frac{\mathrm{d}^+}{\mathrm{d}t} \left(d^2(S[\pi](t), \rho) \right) |_{t=0} + \frac{1}{2} \left| \partial_\pi \left(\frac{1}{2} a d^2(\pi, \rho) \right) \right|^2$$

Applying *EVI* we get a proper upper bound

$$Hf_1(\pi) \le a \left[\mathcal{E}(\rho) - \mathcal{E}(\pi) \right] - a \frac{\kappa}{2} d^2(\pi, \rho) + \frac{1}{2} a^2 d^2(\pi, \rho)$$

as soon as $\mathcal{E}(\rho) < +\infty$

$\Longrightarrow$ define a new Hamiltonian $H_{\mathrm{can},\dagger}$:

$$\mathcal{D}(H_{\mathrm{can},\dagger}) := \left\{ f_1 : E \to \mathbb{R},\ f_1(\pi) = \frac{1}{2} a d^2(\pi, \rho) \,\middle|\, a > 0,\ \rho \in E : \mathcal{E}(\rho) < \infty \right\}$$

and for any $f_1 \in \mathcal{D}(H_{\mathrm{can},\dagger})$ we set

$$H_{\mathrm{can},\dagger} f_1(\pi) = a \left[\mathcal{E}(\rho) - \mathcal{E}(\pi) \right] - a \frac{\kappa}{2} d^2(\pi, \rho) + \frac{1}{2} a^2 d^2(\pi, \rho)$$

IDEA: if $f_1(\pi) = \frac{1}{2} a d^2(\pi, \rho)$ for some $\rho \in E$ and $a > 0$, then formally

$$Hf_1(\pi) = \frac{1}{2} a \frac{\mathrm{d}^+}{\mathrm{d}t} \left(d^2(S[\pi](t), \rho) \right) |_{t=0} + \frac{1}{2} a^2 d^2(\pi, \rho)$$

Applying *EVI* we get a proper upper bound

$$Hf_1(\pi) \leq a \left[\mathcal{E}(\rho) - \mathcal{E}(\pi) \right] - a \frac{\kappa}{2} d^2(\pi, \rho) + \frac{1}{2} a^2 d^2(\pi, \rho)$$

as soon as $\mathcal{E}(\rho) < +\infty$

$\Longrightarrow$ define a new Hamiltonian $H_{\mathrm{can},\dagger}$:

$$\mathcal{D}(H_{\mathrm{can},\dagger}) := \left\{ f_1 : E \to \mathbb{R},\ f_1(\pi) = \frac{1}{2} a d^2(\pi, \rho) \,\middle|\, a > 0,\ \rho \in E : \mathcal{E}(\rho) < \infty \right\}$$

and for any $f_1 \in \mathcal{D}(H_{\mathrm{can},\dagger})$ we set

$$H_{\mathrm{can},\dagger} f_1(\pi) = a \left[\mathcal{E}(\rho) - \mathcal{E}(\pi) \right] - a \frac{\kappa}{2} d^2(\pi, \rho) + \frac{1}{2} a^2 d^2(\pi, \rho)$$

IDEA: if $f_1(\pi) = \frac{1}{2} a d^2(\pi, \rho)$ for some $\rho \in E$ and $a > 0$, then formally

$$Hf_1(\pi) = \frac{1}{2} a \frac{\mathrm{d}^+}{\mathrm{d}t} \left(d^2(S[\pi](t), \rho) \right) |_{t=0} + \frac{1}{2} a^2 d^2(\pi, \rho)$$

Applying *EVI* we get a proper upper bound

$$Hf_1(\pi) \leq a \left[\mathcal{E}(\rho) - \mathcal{E}(\pi) \right] - a \frac{\kappa}{2} d^2(\pi, \rho) + \frac{1}{2} a^2 d^2(\pi, \rho)$$

as soon as $\mathcal{E}(\rho) < +\infty$

$\Longrightarrow$ define a new Hamiltonian $H_{\mathrm{can},\dagger}$:

$$\mathcal{D}(H_{\mathrm{can},\dagger}) := \left\{ f_1 : E \to \mathbb{R},\ f_1(\pi) = \frac{1}{2} a d^2(\pi, \rho) \,\middle|\, a > 0,\ \rho \in E : \mathcal{E}(\rho) < \infty \right\}$$

and for any $f_1 \in \mathcal{D}(H_{\mathrm{can},\dagger})$ we set

$$H_{\mathrm{can},\dagger} f_1(\pi) = a \left[\mathcal{E}(\rho) - \mathcal{E}(\pi) \right] - a \frac{\kappa}{2} d^2(\pi, \rho) + \frac{1}{2} a^2 d^2(\pi, \rho)$$

Similarly, we get a formal lower bound. Let

$$\mathcal{D}(H_{\mathrm{can},\ddagger}) := \left\{ f_2 : E \to \mathbb{R},\ f_2(\mu) = -\frac{1}{2} a d^2(\gamma,\mu) \,\middle|\, a > 0, \gamma \in E : \mathcal{E}(\gamma) < \infty \right\}$$

be the corresponding domain, then for $f_2 \in \mathcal{D}(H_{\mathrm{can},\ddagger})$ we set

$$H_{\mathrm{can},\ddagger} f_2(\mu) = a\left[\mathcal{E}(\mu) - \mathcal{E}(\gamma)\right] + a\frac{\kappa}{2} d^2(\gamma,\mu) + \frac{1}{2} a^2 d^2(\gamma,\mu)$$

and at least formally

$$H_{\mathrm{can},\ddagger} f \leq H f \leq H_{\mathrm{can},\dagger} f$$

Thus, instead of establishing the comparison principle for the original equation we aim to show it for the upper and lower bound we found for our Hamiltonian

The comparison principle in the finite dimensional case

Let u be a subsolution and v a supersolution, we have to prove that

$$\sup_{\pi} u(\pi) - v(\pi) \leq \sup_{\pi} h_1(\pi) - h_2(\pi)$$

IDEA: use a "doubling" variables method

$$\sup_{x \in \mathbb{R}^n} u(x) - v(x) \leq \liminf_{\varepsilon \to 0} \sup_{x,y \in \mathbb{R}^n} u(x) - v(y) - \frac{|x-y|^2}{\varepsilon^2} - \varepsilon(|x|^2 + |y^2|),$$

and observe that

$$(x,y) \mapsto u(x) - v(y) - \frac{|x-y|^2}{\varepsilon^2} - \varepsilon(|x|^2 + |y^2|)$$

admits a global maximum $(x_\varepsilon, y_\varepsilon)$, hence

$$f_1(x) = \frac{|x - y_\varepsilon|^2}{\varepsilon^2} + \varepsilon(|x|^2 + |y_\varepsilon|^2) + v(y_\varepsilon)$$

can be used as a test function in the definition of subsolution to obtain

$$u(x_\varepsilon) - \lambda H f_1(x_\varepsilon) - h_1(x_\varepsilon) \leq 0$$

The comparison principle in the finite dimensional case

Let u be a subsolution and v a supersolution, we have to prove that

$$\sup_{\pi} u(\pi) - v(\pi) \leq \sup_{\pi} h_1(\pi) - h_2(\pi)$$

IDEA: use a "doubling" variables method

$$\sup_{x\in\mathbb{R}^n} u(x) - v(x) \leq \liminf_{\varepsilon\to 0} \sup_{x,y\in\mathbb{R}^n} u(x) - v(y) - \frac{|x-y|^2}{\varepsilon^2} - \varepsilon(|x|^2 + |y^2|),$$

and observe that

$$(x,y) \mapsto u(x) - v(y) - \frac{|x-y|^2}{\varepsilon^2} - \varepsilon(|x|^2 + |y^2|)$$

admits a global maximum $(x_\varepsilon, y_\varepsilon)$, hence

$$f_2(y) = -\frac{|x_\varepsilon - y|^2}{\varepsilon^2} - \varepsilon(|x_\varepsilon|^2 + |y|^2) + u(x_\varepsilon)$$

can be used as a test function in the definition of supersolution to obtain

$$v(y_\varepsilon) - \lambda H f_2(y_\varepsilon) - h_2(y_\varepsilon) \geq 0$$

The comparison principle in the finite dimensional case

Let u be a subsolution and v a supersolution, we have to prove that

$$\sup_{\pi} u(\pi) - v(\pi) \leq \sup_{\pi} h_1(\pi) - h_2(\pi)$$

IDEA: use a "doubling" variables method

$$\sup_{x\in\mathbb{R}^n} u(x) - v(x) \leq \liminf_{\varepsilon\to 0} \sup_{x,y\in\mathbb{R}^n} u(x) - v(y) - \frac{|x-y|^2}{\varepsilon^2} - \varepsilon(|x|^2 + |y^2|),$$

and observe that

$$(x,y) \mapsto u(x) - v(y) - \frac{|x-y|^2}{\varepsilon^2} - \varepsilon(|x|^2 + |y^2|)$$

admits a global maximum $(x_\varepsilon, y_\varepsilon)$, hence

$$\sup_{x\in\mathbb{R}^n} u(x) - v(x) \leq \liminf_{\varepsilon\to 0} \lambda(Hf_1(x_\varepsilon) - Hf_2(y_\varepsilon)) + h_1(x_\varepsilon) - h_2(y_\varepsilon)$$

$$\leq \sup_{x\in\mathbb{R}^n} h_1(x) - h_2(x)$$

The comparison principle

Let u be a subsolution and v a supersolution, we have to prove that

$$\sup_{\pi} u(\pi) - v(\pi) \leq \sup_{\pi} h_1(\pi) - h_2(\pi)$$

The key steps :

- perform doubling variables method
- prove the existence of a global maximum $\rightsquigarrow$ Ekeland's perturbed optimization principle
- use of sub(super)solutions properties
- estimate of the difference of the Hamiltonians

Coming back to our case: We perform the classical "doubling" variables method using the distance function and the energy functional

$$\sup_{\pi \in E} u(\pi) - v(\pi) \leq \liminf_{\varepsilon \to 0} \sup_{\pi, \mu \in E} \mathcal{G}_\varepsilon(\pi, \mu),$$

where $\mathcal{G}_\varepsilon : E^2 \to [-\infty, +\infty)$,

$$\mathcal{G}_\varepsilon(\pi, \mu) := u(\pi) - v(\mu) - \frac{1}{\varepsilon^2} d^2(\pi, \mu) - \varepsilon(\mathcal{E}(\pi) + \mathcal{E}(\mu))$$

PROBLEM 1: $\mathcal{E}$ is not bounded below, however, as we have seen, as a consequence of EVI, $\mathcal{E}$ can bounded from below by a non-negative constant times $-d^2$ $\rightsquigarrow$ we use $\overline{\mathcal{E}}$

PROBLEM 2: $\mathcal{E}(\pi), \mathcal{E}(\mu)$ can be $+\infty$, hence we cannot leave them as a free variable in the test function $\rightsquigarrow$ we have to quadruplicate variables

We perform quadruplication of variables and use a different $\mathcal{G}_{\alpha,\varepsilon}$

$$\sup_{\pi \in E} u(\pi) - v(\pi) \leq \liminf_{\varepsilon \to 0} \sup_{\pi,\mu \in E \rho,\gamma \in \mathcal{D}(\mathcal{E})} \mathcal{G}_{\alpha,\varepsilon}(\pi,\mu,\rho,\gamma),$$

where, for given $\alpha > 0$, $\mathcal{G}_{\alpha,\varepsilon} : E^4 \to [-\infty,+\infty)$,

$$\begin{aligned} \mathcal{G}_{\alpha,\varepsilon}(\pi,\mu,\rho,\gamma) := & \frac{u(\pi)}{1-\varepsilon} - \frac{v(\mu)}{1+\varepsilon} \\ & - \frac{\alpha}{2(1-\varepsilon)} d^2(\rho,\pi) - \frac{\alpha}{2} d^2(\rho,\gamma) - \frac{\alpha}{2(1+\varepsilon)} d^2(\gamma,\mu) \\ & - \frac{\varepsilon}{1-\varepsilon} \overline{\mathcal{E}}(\rho) - \frac{\varepsilon}{1+\varepsilon} \overline{\mathcal{E}}(\gamma) \end{aligned}$$

NOTE: This procedure is actually reminiscent of the sup-convolution procedure

$$\sup_{\pi\in E} u(\pi) - v(\pi) \le \liminf_{\varepsilon\to 0} \sup_{\pi,\mu\in E\,\rho,\gamma\in\mathcal{D}(\mathcal{E})} \mathcal{G}_{\alpha,\varepsilon}(\pi,\mu,\rho,\gamma),$$

As it is usually done for infinite dimensional problems, to obtain optimizers we have to use Ekeland's perturbed optimization principle with an appropriate penalization function

Indeed Ekeland's perturbed optimization principle claims that, if we add a small perturbation to the test function, we can always attain the extremum

PROBLEM: Choose a useful penalization function to obtain good estimates for the difference of the Hamiltonians $\rightsquigarrow$ use Tataru distance

Note: If we modify the test functions $\rightsquigarrow$ we have to define a new Hamiltonian $H_\dagger$

IDEA: The classical Tataru distance was defined on Hilbert spaces for contracting gradient flows as

$$d_T(\mu,\nu) := \inf_{t\geq 0} \{t + d(\mu, S[\nu](t))\}$$

- d_T is not a metric because it is not symmetric
- d_T is Lipschitz with respect to the metric d
- d_T behaves well with respect to the corresponding gradient flow

$$\frac{d_T(S[\nu](r),\hat{\nu}) - d_T(\nu,\hat{\nu})}{r} \leq 1$$

for all $\forall \nu, \hat{\nu} \in E,\ r \in \mathbb{R} \setminus \{0\}$

PROBLEM: Our gradient flow is not necessarily contracting with respect to the metric (a negative κ is allowed) $\rightsquigarrow$ we have to work with an adjusted Tataru distance

The right generalization of Tataru distance turns out to be defined on $E \times E$ as

$$d_T(\mu,\nu) := \inf_{t\geq 0}\left\{t + e^{\hat{\kappa}t} d(\mu, S[\nu](t))\right\}$$

where $\hat{\kappa} = 0 \wedge \kappa \leq 0$

We recover the same key properties of the classical Tataru distance

- In particular, we have

$$\left|\frac{d^+}{dt}(d_T(S[\pi](t),\mu))|_{t=0}\right| \leq 1$$

  and

$$|\partial_\pi d_T(\pi,\mu)| \leq 1$$

- d_T allows us remove compactness assumptions both for the level sets of the energy functional $\mathcal{E}$ and for metric balls we use new test functions

$$f_1(\pi) = \frac{1}{2} a d^2(\pi, \rho) + b d_T(\pi, \mu) + c$$

for $a, b > 0$, $c \in \mathbb{R}$, and $\rho, \mu \in E$ such that $\mathcal{E}(\rho) < \infty$

$$f_2(\mu) = -\frac{1}{2} a d^2(\gamma, \mu) - b d_T(\mu, \pi) + c$$

for $a, b > 0$, $c \in \mathbb{R}$, and $\gamma, \pi \in E$ such that $\mathcal{E}(\gamma) < \infty$

and using the same strategy we modify the Hamiltonians in this way

$$H_\dagger f_1(\pi) = a\left[\mathcal{E}(\rho) - \mathcal{E}(\pi)\right] - a\frac{\kappa}{2} d^2(\pi, \rho) + b + \frac{1}{2} a^2 d^2(\pi, \rho) + abd(\pi, \rho) + \frac{1}{2} b^2$$

$$H_\ddagger f_2(\mu) = a\left[\mathcal{E}(\mu) - \mathcal{E}(\gamma)\right] + a\frac{\kappa}{2} d^2(\gamma, \mu) - b + \frac{1}{2} a^2 d^2(\gamma, \mu) - abd(\gamma, \mu)$$

As a penalization function, we use $\mathcal{B}_\varepsilon : E^4 \times E^4 \to [0,+\infty)$ defined as

$$\mathcal{B}_\varepsilon(\pi,\mu,\rho,\gamma,\tilde\pi,\tilde\mu,\tilde\rho,\tilde\gamma) := \frac{1}{1-\varepsilon} d_T(\pi,\tilde\pi) + \frac{1}{1+\varepsilon} d_T(\mu,\tilde\mu) + d_T(\rho,\tilde\rho) + d_T(\gamma,\tilde\gamma)$$

then $\mathcal{B}_\varepsilon$ and $\mathcal{G}_{\alpha,\varepsilon}$ satisfy the hypotheses of Ekeland's principle

i.e. $\exists !\ x_\alpha = (\pi_\alpha,\mu_\alpha,\rho_\alpha,\gamma_\alpha) \in E^2 \times (\mathcal{D}(\mathcal{E}))^2$ s.t.

$$\mathcal{G}_{\alpha,\varepsilon}(x_\alpha) = \sup_{\pi,\mu\in E\,\rho,\gamma\in\mathcal{D}(\mathcal{E})} \mathcal{G}_{\alpha,\varepsilon}(\pi,\mu,\rho,\gamma) - \frac{1}{\alpha}\mathcal{B}_\varepsilon(\pi,\mu,\rho,\gamma,x_\alpha)$$

and

$$\sup_{\pi\in E} u(\pi) - v(\pi) \le \liminf_{\alpha\to+\infty} \liminf_{\varepsilon\to 0} \sup_{\pi,\mu\in E\,\rho,\gamma\in\mathcal{D}(\mathcal{E})} \mathcal{G}_{\alpha,\varepsilon}(\pi,\mu,\rho,\gamma) - \frac{1}{\alpha}\mathcal{B}_\varepsilon(\pi,\mu,\rho,\gamma,x_\alpha)$$

Defining test functions f_1, f_2 so that

$$u(\pi) - f_1(\pi) = (1-\varepsilon)[\mathcal{G}_{\alpha,\varepsilon} - \frac{1}{\alpha}\mathcal{B}_\varepsilon(\cdot, x_\alpha)](\pi, \mu_\alpha, \rho_\alpha, \gamma_\alpha),$$

and

$$v(\mu) - f_2(\mu) = -(1+\varepsilon)[\mathcal{G}_{\alpha,\varepsilon} - \frac{1}{\alpha}\mathcal{B}_\varepsilon(\cdot, x_\alpha)](\pi_\alpha, \mu, \rho_\alpha, \gamma_\alpha),$$

we obtain, for the right choice of $\varepsilon = \varepsilon_\alpha \to 0$ as $\alpha \to +\infty$,

$$\sup_{\pi \in E} u(\pi) - v(\pi) \leq \frac{h_1(\pi_\alpha)}{1-\varepsilon_\alpha} - \frac{h_2(\mu_\alpha)}{1+\varepsilon_\alpha} + \lambda\Big(\frac{1}{1-\varepsilon_\alpha} H_\dagger f_1(\pi_\alpha) - \frac{1}{1+\varepsilon_\alpha} H_\ddagger f_2(\mu_\alpha)\Big) + \mathcal{O}(\alpha^{-1/2}).$$

useful estimates for the difference of the Hamiltonians are recovered using *EVI* inequality and the properties of the gradient flow

Comparison with the master equation

The master equation, arising in the theory of mean field games, shares some properties with infinite dimensional HJ equations characterizing the value function of McKean-Vlasov control problems, see Bensoussan, Frehse and Yam 2013

As explained in Carmona, Delarue, and Lachapelle 2013 Mean Field games and McKean-Vlasov control problems remain conceptually different, essentially the two methods differ in the order in which optimization and passage to the limit are performed

Comparing the monotonicity assumption that is typically imposed on the coefficients of the master equation and the geodesic convexity of the energy functional that underlies our computations, these two geometrical assumptions are not directly related and enter the respective equations in a different way

Gangbo and Mészáros 2020 use displacement convexity (in a different way than we do) to prove well-posedness of potential master equations

The Schrödinger problem

The Schrödinger problem (SP) is the problem of finding the most likely evolution of a cloud of independent Brownian particles conditionally on the observation of their initial and final configuration,
i.e. an entropy minimization problem with marginal constraints

SP is the object of a very dynamic research activity:

It has powerful connections with the theory of Large Deviations, PDEs, Optimal transport, statistical machine learning and numerical algorithms for PDE related problems

KEY IDEA: SP may be viewed as a (entropic) regularization of the Optimal Transport problem

The Mean Field Schrödinger problem

The Mean Field Schrödinger Problem (MFSP) is obtained by replacing in the previous description the independent particles by interacting ones

Interacting Particle System

$(\Omega, \mathcal{F}_t, \mathcal{F}_T)$ where $\Omega = C([0,T];\mathbb{R}^d)$ with the uniform topology and $\{\mathcal{F}_t\}_{t\in[0,T]}$ the coordinate filtration

Interaction Potential: a symmetric $\mathcal{C}^2$ function $W:\mathbb{R}^d \to \mathbb{R}$ s.t. $\sup_{z,v\in\mathbb{R}^d,|v|=1} v\cdot\nabla^2 W(z)\cdot v < +\infty$

For N large, we consider Brownian particles $(X_t^{i,N})_{t\in[0,T],1\leq i\leq N}$

$$\begin{cases} \mathrm{d}X_t^{i,N} = & -\frac{1}{N}\sum_{k=1}^N \nabla W(X_t^{i,N}-X_t^{k,N})\mathrm{d}t + \mathrm{d}B_t^i \\ X_0^{i,N} \sim & \mu^{\mathrm{in}} \in \mathcal{P}_2(\mathbb{R}^d) \end{cases}$$

Driving Question: If at time T we observe that the sequence of empirical path measures

$$\frac{1}{N}\sum_{i=1}^N \delta_{X_T^{i,N}} \approx \mu^{\mathrm{fin}} \in \mathcal{P}_2(\mathbb{R}^d),$$

what have done the particles in between?

Denote by

$$\Pi(\mu^{\mathrm{in}}, \mu^{\mathrm{fin}}) := \left\{P \in \mathcal{P}_1(C([0,T];\mathbb{R}^d)) : P_0 = \mu^{\mathrm{in}}, P_T = \mu^{\mathrm{fin}}\right\}$$

and for $P, Q \in \mathcal{P}_1(C([0,T];\mathbb{R}^d))$, let $\mathcal{H}(P|Q)$ denote the relative entropy of P with respect to Q,

$$\mathcal{H}(P|Q) = \begin{cases} \mathbb{E}_P\left[\log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\right] & P \ll Q \\ +\infty & \text{otherwise} \end{cases}$$

$\frac{\mathrm{d}P}{\mathrm{d}Q}$ denotes the Radon-Nikodym density of P against Q

The mean field Schrödinger problem can be stated as

$$\mathscr{C}_T(\mu^{\rm in},\mu^{\rm fin}) := \inf\left\{\mathcal{H}(P|\Gamma(P)) \ : \ \ P\in\Pi(\mu^{\rm in},\mu^{\rm fin})\right\}$$

where $\Gamma(P)$ is the law of the unique solution to

$$\begin{cases} \mathrm{d}X_t = & -\nabla W * P_t(X_t)\mathrm{d}t + \mathrm{d}B_t \\ X_0 \sim & \mu^{\rm in} \end{cases}$$

Its optimal value is called mean field entropic transportation cost and its optimizers are called mean field Schrödinger bridges (MFSB)

Theorem (Backhoff, Conforti, Gentil, Léonard '19)

Under mild assumptions MFSB exist

Uniqueness is still an open question

Connections with MFG

Theorem (BCGL '19)

Let P be an optimizer for (MFSP). Then there exists a weak gradient field Ψ s.t.

$$\mathrm{d}X_t = (\Psi_t(X_t) - \nabla W * P_t(X_t))\mathrm{d}t + dB_t$$

Now, set $\mu_t = (X_t)_{\#}P$ for all $t \in [0, T]$ and let μ and Ψ be $\mathcal{C}^{1,2}$, $\mu > 0$
Then there exists $\psi : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\Psi_t(x) = \nabla\psi_t(x) \quad \forall t \in [0, T], x \in \mathbb{R}^d$$

and $(\psi(\cdot), \mu(\cdot))$ is a classical solution of the following mean field planning PDE system

$$\begin{cases} \partial_t\psi_t(x) + \frac{1}{2}\Delta\psi_t(x) + \frac{1}{2}\left|\nabla\psi_t(x)\right|^2 = \int_{\mathbb{R}^d}\nabla W(x-\tilde{x})\cdot(\nabla\psi_t(x) - \nabla\psi_t(\tilde{x}))\,\mu_t(\mathrm{d}\tilde{x}) \\ \partial_t\mu_t(x) - \frac{1}{2}\Delta\mu_t(x) + \nabla\cdot((-\nabla W * \mu_t(x) + \nabla\psi_t(x))\,\mu_t(x)) = 0 \\ \mu_0(x) = \mu^{\mathrm{in}}(x), \mu_T(x) = \mu^{\mathrm{fin}}(x) \end{cases}$$

This type of PDE system has a similar structure to the planning MFG

The HJ equation on the space of probability measures

Mimicking the well-known duality between the Monge-Kantorovich problem and the Hamilton-Jacobi equation, the MFSP can be formally seen as in duality with the solution of an infinite dimensional Hamilton-Jacobi (HJ) equation in $\mathcal{P}_2(\mathbb{R}^d)$

Let us modify the problem adding a penalization at the final time and removing the corresponding marginal constraint

For all $t \in [0, T]$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ we define

$$u(t,\mu) := \inf\{\mathcal{H}(P|\Gamma(P)) + \mathcal{G}(P_T) : P \in \mathcal{P}_2(\Omega), P_t = \mu\}$$

As for the classical MFSP, the previous problem can be rewritten equivalently as

$$u(t,\mu) := \inf \frac{1}{2}\int_t^T\int_{\mathbb{R}^d}\left|w_s(z) + \frac{1}{2}\nabla\log\mu_s(z) + \nabla W * \mu_s(z)\right|^2 \mu_s(\mathrm{d}z)\mathrm{d}s + \mathcal{G}(\mu_T)$$

over all absolutely continuous curves $(\mu_s)_{s\in[t,T]} \subset \mathcal{P}_2(\mathbb{R}^d)$ s.t. that are weak solutions of the following continuity equation

$$\partial_s\mu_s + \nabla\cdot(w\mu_s) = 0 \quad \mu_t = \mu,$$

Then the optimal value $u(t,\mu)$ is a candidate solution for an HJ equation on the space of probability measures

The HJ equation on the space of probability measures

Formally the HJ equation looks like

$$\begin{cases} -\partial_t u(t,\mu) + Hu(t,\mu) = 0, \\ u(T,\mu) = \mathcal{G}(\mu_T) \end{cases}$$

where the Hamiltonian is written as an operator over functions on $\mathcal{P}_2(\mathbb{R}^d)$

$$Hu(\mu) = \frac{1}{2}\int_{\mathbb{R}^d} \langle \mathrm{grad}^{W_2} u(\mu), \mathrm{grad}^{W_2} \tilde{\mathcal{F}}(\mu)\rangle \mu(\mathrm{d}x) + \frac{1}{2}\int_{\mathbb{R}^d} |\mathrm{grad}^{W_2} u(\mu)|^2 \mu(\mathrm{d}x)$$

where the free energy functional is defined for $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ as

$$\mu \mapsto \tilde{\mathcal{F}}(\mu) := \begin{cases} \int \log \mu(x)\mu(\mathrm{d}x) + \int\int W(x-y)\mu(\mathrm{d}y)\mu(\mathrm{d}x) & \mu \ll \mathcal{L} \\ +\infty & \text{otherwise} \end{cases}$$

Existence of Solutions

Let (E, d) be $(\mathcal{P}_2(\mathbb{R}^d), W_2)$

IDEA: show that

$$f(\rho) = \sup\left\{\int_0^{+\infty} e^{-\lambda^{-1}t}\left[\lambda^{-1}h(\rho^u(t)) - \frac{1}{2}\|u(t)\|^2_{L^2\rho^u(t)}\right]\mathrm{d}t\right\},$$

where the sup is taken over all $(\rho^u(\cdot), u(\cdot))$ weak solution of

$$\partial_t\rho^u - \frac{1}{2}\Delta\rho^u + \nabla\cdot(\rho^u u) = 0, \quad \rho^u(0) = \rho,$$

is a viscosity solution of

$$f - \lambda Hf = h,$$

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

where $\mathcal{E}(\mu) = \int \mu \log \mu$

Definition of viscosity super and sub solutions

Given the stationary HJ equation

$$f - \lambda H f = h$$

Definition: We say that $u : E \to \mathbb{R}$ is a (viscosity) subsolution if u is bounded, upper semi-continuous and if for all $f_1 \in \mathcal{D}(H)$ and $\rho_0 \in E$ s.t.

$$u(\rho_0) - f_1(\rho_0) = \sup_{\rho} u(\rho) - f_1(\rho)$$

we have

$$u(\rho_0) - \lambda H f_1(\rho_0) - h(\rho_0) \leq 0$$

Definition of viscosity super and sub solutions

Given the stationary HJ equation

$$f - \lambda H f = h$$

Definition: We say that $v : E \to \mathbb{R}$ is a (viscosity) supersolution if v is bounded, lower semi-continuous and if for all $f_2 \in \mathcal{D}(H)$ and $\rho_0 \in E$ s.t.

$$v(\rho_0) - f_2(\rho_0) = \inf_{\rho} v(\rho) - f_2(\rho),$$

we have

$$v(\rho_0) - \lambda H f_2(\rho_0) - h(\rho_0) \geq 0$$

Ingredients:

- use smooth test functions (the Tataru distance is not smooth)
- relax the control problem to prove weak regularity of the value function
- use the Dynamic Programming Principle
- use a modified *EVI* inequality satisfied by a the controlled gradient flow

Future directions

Our Aims are:

- conclude the result on Existence of solutions for HJ
- Time dependent problem
- Long time behavior
- study a richer class of equations, possibly including a stochastic component modeling a source of common noise