Some "geometric" tools to study Hamilton-Jacobi-Bellman equations on the Wasserstein space

Charles Bertucci, CNRS, CEREMADE, Université Paris Dauphine-PSL.

20/11/2025, Orsay, Geometry, duality and convexity in new OT problems

HJB equations on the space of probability measures

- The central object of this talk is HJB equations of the form

$$-\partial_t U(t,\mu) + \int_{\mathbb{R}^d} H(x, D_\mu U(t,\mu)(x), \mu)\mu(dx) - \sigma A[U,\mu] = 0$$
$$\text{in } (0,T) \times \mathcal{P}_2(\mathbb{R}^d),$$
$$U|_{t=T} = G.$$

- For most of the talk $\sigma = 0$...
- Main questions are of existence, uniqueness, stability of solutions
- Main challenge is that typical solutions are not smooth...

Motivations and main challenges

Main notation and concepts

- We will work on $\mathcal{P}_2(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d), \int_{\mathbb{R}^d} |x|^2 \mu(dx) < \infty\}$.

Main notation and concepts

- We will work on $\mathcal{P}_2(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d), \int_{\mathbb{R}^d} |x|^2\mu(dx) < \infty\}$.
- We endow it with the 2-Wasserstein distance

$$W_2(\mu,\nu) = \min_{\gamma\in\Pi(\mu,\nu)} \left(\int_{\mathbb{R}^{2d}} |x-y|^2\gamma(dx,dy)\right)^{\frac{1}{2}}.$$

Main notation and concepts

- We will work on $\mathcal{P}_2(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d), \int_{\mathbb{R}^d} |x|^2 \mu(dx) < \infty\}$.
- We endow it with the 2-Wasserstein distance

$$W_2(\mu,\nu) = \min_{\gamma \in \Pi(\mu,\nu)} \left(\int_{\mathbb{R}^{2d}} |x-y|^2 \gamma(dx,dy) \right)^{\frac{1}{2}}.$$

- We will differentiate $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ according to

$$\lim_{t\to 0} \frac{U(m_t) - U(m_0)}{t} = \int_{\mathbb{R}^d} D_\mu U(m_0)(x) \cdot \phi(x) m_0(dx),$$

where

$$\partial_t m + \operatorname{div}(\phi m) = 0 \text{ in } (-T,T) \times \mathbb{R}^d,$$
$$m|_{t=0} = m_0,$$

for some $\phi : \mathbb{R}^d \to \mathbb{R}^d$.

Optimal control of measures

- Starting from a probability measure $\mu \in \mathcal{P}(\Omega)$ [1] at time $t = 0$, consider the problem of controlling an evolution $(m_t)_{t\in[0,T]}$ so that to minimize a certain cost.

1. Ω is either $\mathbb{T}^d$ or $\mathbb{R}^d$ here.

Optimal control of measures

- Starting from a probability measure $\mu \in \mathcal{P}(\Omega)$ [1] at time $t = 0$, consider the problem of controlling an evolution $(m_t)_{t\in[0,T]}$ so that to minimize a certain cost.
- E.g. dynamics are given by

$$\partial_t m + \mathrm{div}(\alpha m) = 0 \text{ in } (0, T) \times \Omega,$$

where $\alpha : [0, T] \times \Omega \to \mathbb{R}^d$ is the control chosen.

1. Ω is either $\mathbb{T}^d$ or $\mathbb{R}^d$ here.

Optimal control of measures

- Starting from a probability measure $\mu \in \mathcal{P}(\Omega)$ [1] at time $t = 0$, consider the problem of controlling an evolution $(m_t)_{t\in[0,T]}$ so that to minimize a certain cost.
- E.g. dynamics are given by

$$\partial_t m + \mathrm{div}(\alpha m) = 0 \text{ in } (0, T) \times \Omega,$$

where $\alpha : [0, T] \times \Omega \to \mathbb{R}^d$ is the control chosen.
- Typical minimization problem are given by

$$\inf_{\alpha,m} \int_0^T \int_\Omega L(x, \alpha(t, x), m_t) m_t(dx) dt + G(m_T).$$

1. Ω is either $\mathbb{T}^d$ or $\mathbb{R}^d$ here.

Optimal control of measures

- Starting from a probability measure $\mu \in \mathcal{P}(\Omega)$ [1] at time $t = 0$, consider the problem of controlling an evolution $(m_t)_{t\in[0,T]}$ so that to minimize a certain cost.
- E.g. dynamics are given by

$$\partial_t m + \text{div}(\alpha m) = 0 \text{ in } (0, T) \times \Omega,$$

where $\alpha : [0, T] \times \Omega \to \mathbb{R}^d$ is the control chosen.
- Typical minimization problem are given by

$$\inf_{\alpha, m} \int_0^T \int_\Omega L(x, \alpha(t, x), m_t) m_t(dx) dt + G(m_T).$$

1. Ω is either $\mathbb{T}^d$ or $\mathbb{R}^d$ here.

Optimal control of measures

- Starting from a probability measure $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ at time $t = 0$, consider the problem of controlling an evolution $(m_t)_{t\in[0,T]}$ so that to minimize a certain cost.
- E.g. dynamics are given by

$$\partial_t m + \mathrm{div}(\alpha m) - \sigma\Delta m = 0 \text{ in } (0, T) \times \Omega,$$

where $\alpha : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ is the control chosen.
- Typical minimization problem are given by

$$\inf_{\alpha, m} \int_0^T \int_{\mathbb{R}^d} L(x, \alpha(t, x), m_t) m_t(dx) dt + G(m_T).$$

Optimal control of measures II

- Using Bellman's dynamic programming, we introduce the value function U defined by

$$U(t,\mu) = \inf_{\alpha,m} \int_t^T \int_{\mathbb{R}^d} L(x,\alpha_s(x),m_s)m_s(dx)ds + G(m_T).$$

where the infimum is on paths m starting from μ at time t.

Optimal control of measures II

- Using Bellman's dynamic programming, we introduce the value function U defined by

$$U(t,\mu) = \inf_{\alpha,m} \int_t^T \int_{\mathbb{R}^d} L(x,\alpha_s(x),m_s)m_s(dx)ds + G(m_T).$$

where the infimum is on paths m starting from μ at time t.

- Formally, U solves

$$-\partial_t U(t,\mu) + \int_\Omega H(x, D_\mu U(t,\mu)(x),\mu)\mu(dx) = 0 \text{ in } (0,T)\times\mathcal{P}(\Omega),$$

where $H(x,p,\mu) = \sup_\alpha\{-L(x,\alpha,\mu) - \alpha\cdot p\}$.

Derivation of the HJB equation

- Formally

$$U(t,\mu) = \inf_{(\alpha,m)} \left\{ \int_t^{t+\kappa} \int_\Omega L(x,\alpha,m_s)dmds + U(t+\kappa, m(t+\kappa)) \right\}$$

$$0 = \inf_{(\alpha,m)} \left\{ \int_t^{t+\kappa} \int_\Omega L(x,\alpha,m)dmds + \partial_t U(t,\mu)\kappa + o(\kappa) + \right.$$
$$\left. + \int_t^{t+\kappa} \int_\Omega D_\mu U(s,m_s)\alpha_s(x)dm_s ds \right\}$$

Derivation of the HJB equation

- Formally

$$U(t,\mu) = \inf_{(\alpha,m)} \left\{ \int_t^{t+\kappa} \int_\Omega L(x,\alpha,m_s)dmds + U(t+\kappa, m(t+\kappa)) \right\}$$

$$0 = \inf_{(\alpha,m)} \left\{ \int_t^{t+\kappa} \int_\Omega L(x,\alpha,m)dmds + \partial_t U(t,\mu)\kappa + o(\kappa) + \right.$$
$$\left. + \int_t^{t+\kappa} \int_\Omega D_\mu U(s,m_s)\alpha_s(x)dm_s ds \right\}$$

- Dividing by κ and taking the limit $\kappa \to 0$, we obtain

$$-\partial_t U - \inf_\alpha \left\{ \int_\Omega (L(x,\alpha,\mu) + D_\mu U(t,\mu) \cdot \alpha) d\mu \right\} = 0.$$

Derivation of the HJB equation

- Formally

$$U(t,\mu) = \inf_{(\alpha,m)} \left\{ \int_t^{t+\kappa} \int_\Omega L(x,\alpha,m_s) dm ds + U(t+\kappa, m(t+\kappa)) \right\}$$

$$0 = \inf_{(\alpha,m)} \left\{ \int_t^{t+\kappa} \int_\Omega L(x,\alpha,m) dm ds + \partial_t U(t,\mu)\kappa + o(\kappa) + \right.$$
$$\left. + \int_t^{t+\kappa} \int_\Omega D_\mu U(s,m_s)\alpha_s(x) dm_s ds \right\}$$

- Dividing by κ and taking the limit $\kappa \to 0$, we obtain

$$-\partial_t U - \inf_\alpha \left\{ \int_\Omega (L(x,\alpha,\mu) + D_\mu U(t,\mu)\cdot\alpha) d\mu \right\} = 0.$$

- Denoting $H(x,p,m) = \sup_\alpha \{-L(x,\alpha,m) - \alpha\cdot p\}$, we arrive at the equation

$$-\partial_t U + \int_\Omega H(x, D_\mu U(t,\mu)(x), \mu)\mu(dx) = 0 \text{ in } (0,T)\times\mathcal{P}(\mathbb{T}^d).$$

Typical problem

- Take $G(\mu) = W_2^2(\mu,\nu)$ and $L(x,\alpha,\mu) = |\alpha|^2$.
- HJB equation is

$$-\partial_t U + \frac{1}{4}\int_\Omega |D_\mu U(t,\mu)|^2 d\mu = 0.$$

- The unique solution is

$$U(t,\mu) = \frac{1}{1+T-t} W_2^2(\mu,\nu),$$

which is not differentiable!!

HJB for large deviations

- Typical question in Large deviation : compute quantity such as $V(t,x) := \mathbb{P}(X_T \in A|\ X_t = x)$.

HJB for large deviations

- Typical question in Large deviation : compute quantity such as $V(t,x) := \mathbb{P}(X_T \in A|\ X_t = x)$.
- Feynmac-Kac formulas implies that V solves a linear equation.

HJB for large deviations

- Typical question in Large deviation : compute quantity such as $V(t,x) := \mathbb{P}(X_T \in A|\ X_t = x)$.
- Feynmac-Kac formulas implies that V solves a linear equation.
- Heuristics hint to look for the Hopf-Cole transfor, i.e. looking for an equation on $U = -\log(\beta V)$, we end up with a quadratic HJB equation.

HJB for large deviations

- Typical question in Large deviation : compute quantity such as $V(t,x) := \mathbb{P}(X_T \in A|\ X_t = x)$.
- Feynmac-Kac formulas implies that V solves a linear equation.
- Heuristics hint to look for the Hopf-Cole transfor, i.e. looking for an equation on $U = -\log(\beta V)$, we end up with a quadratic HJB equation.
- Quite general idea which holds also in mean field setting.

Why bother with HJB equations ?

- If HJB equations arise from optimal control of measures, why not use the same techniques as in optimal transport ?

Why bother with HJB equations ?

- If HJB equations arise from optimal control of measures, why not use the same techniques as in optimal transport ?
- When the problem is stochastic, the PDE is much more convenient

$$\inf_{\alpha,m} \mathbb{E}\left[\int_0^T \int_\Omega L(x,\alpha(t,x),m_t,p_t)m_t(dx)dt + G(m_T,p_T)\right],$$

where

$$dp_t = b(p_t)dt + \sigma dW_t.$$

- This can model OT with stochastic target e.g.

Main objectives

- When studying PDE of the form

$$-\partial_t U + \mathcal{H}(\mu, D_\mu U) = 0$$

main tool is the one of viscosity solutions.

Main objectives

- When studying PDE of the form

$$-\partial_t U + \mathcal{H}(\mu, D_\mu U) = 0$$

  main tool is the one of viscosity solutions.
- This boils down to understanding how comparison of solutions works, i.e. if $U(T,\mu) \leq V(T,\mu)$ and

$$-\partial_t U + \mathcal{H}(\mu, D_\mu U) \leq 0$$
$$-\partial_t V + \mathcal{H}(\mu, D_\mu V) \geq 0,$$

  then $U \leq V$ everywhere.

Main objectives

- When studying PDE of the form

$$-\partial_t U + \mathcal{H}(\mu, D_\mu U) = 0$$

  main tool is the one of viscosity solutions.
- This boils down to understanding how comparison of solutions works, i.e. if $U(T,\mu) \leq V(T,\mu)$ and

$$-\partial_t U + \mathcal{H}(\mu, D_\mu U) \leq 0$$
$$-\partial_t V + \mathcal{H}(\mu, D_\mu V) \geq 0,$$

  then $U \leq V$ everywhere.
- Rem : Comparison of smooth solution is easy since at points of max $U - V$, then $D_\mu U = D_\mu V$.

Comparison principles and viscosity solutions

The finite dimensional case

- Consider, for $\lambda > 0$, the equation

$$\lambda u + H(x, \nabla_x u) = f(x) \text{ on } \mathbb{T}^d$$

- To compare sub-solution u (ucs) and super-solution v (lsc), consider (x_ϵ, y_ϵ) point of maximum of

$$(x, y) \to u(x) - v(y) - \frac{1}{2\epsilon}(x - y)^2.$$

- Using the viscosity solutions properties

$$\lambda u(x_\epsilon) + H\left(x_\epsilon, \frac{1}{\epsilon}(x_\epsilon - y_\epsilon)\right) \leq f(x_\epsilon),$$
$$\lambda v(y_\epsilon) + H\left(y_\epsilon, \frac{1}{\epsilon}(x_\epsilon - y_\epsilon)\right) \geq f(y_\epsilon).$$

The finite dimensional case II

- Taking the difference, we obtain, noting $p_\epsilon = \epsilon^{-1}(x_\epsilon - y_\epsilon)$,

$$\lambda \max(u-v) \leq \lambda(u(x_\epsilon)-v(y_\epsilon)) \leq f(x_\epsilon)-f(y_\epsilon)+H(y_\epsilon, p_\epsilon)-H(x_\epsilon, p_\epsilon)$$

The finite dimensional case II

- Taking the difference, we obtain, noting $p_\epsilon = \epsilon^{-1}(x_\epsilon - y_\epsilon)$,

$$\lambda \max(u-v) \leq \lambda(u(x_\epsilon)-v(y_\epsilon)) \leq f(x_\epsilon)-f(y_\epsilon)+H(y_\epsilon, p_\epsilon)-H(x_\epsilon, p_\epsilon)$$

- Since $\epsilon^{-1}|x_\epsilon - y_\epsilon|^2 \to_{\epsilon\to 0} 0$, the result follows from assumptions like

$$\begin{cases} f \text{ is continuous,} \\ H(y,p) - H(x,p) \leq C(1+|p|)|x-y|. \end{cases}$$

The finite dimensional case II

- Taking the difference, we obtain, noting $p_\epsilon = \epsilon^{-1}(x_\epsilon - y_\epsilon)$,

$$\lambda \max(u-v) \leq \lambda(u(x_\epsilon)-v(y_\epsilon)) \leq f(x_\epsilon)-f(y_\epsilon)+H(y_\epsilon, p_\epsilon)-H(x_\epsilon, p_\epsilon)$$

- Since $\epsilon^{-1}|x_\epsilon - y_\epsilon|^2 \to_{\epsilon\to 0} 0$, the result follows from assumptions like

$$\begin{cases} f \text{ is continuous,} \\ H(y,p) - H(x,p) \leq C(1+|p|)|x-y|. \end{cases}$$

- Problem ! in infinite dimension such assumptions are not reasonable and the squared distance is not smooth...

$$\mathcal{H}(\mu, \varphi) = \int_\Omega H(\varphi(x))\mu(dx), \text{ in } \mathcal{P}_2(\mathbb{R}^d) \times L^2_\mu.$$

Passing through super-differential and a bit of geometry

- $W_2^2(\cdot, \nu)$ is not differentiable but it is super-differentiable everywhere !

Passing through super-differential and a bit of geometry

- $W_2^2(\cdot,\nu)$ is not differentiable but it is super-differentiable everywhere !
- We have the computation

$$\begin{aligned} W_2^2(\mu',\nu) - W_2^2(\mu,\nu) &\leq \frac{\lambda}{2}\mathbb{E}[|X'-Y|^2] - \frac{\lambda}{2}\mathbb{E}[|X-Y|^2] \\ &= \frac{\lambda}{2}\mathbb{E}[|X'-X+X-Y|^2] - \frac{\lambda}{2}\mathbb{E}[|X-Y|^2] \\ &= \mathbb{E}[\lambda(X-Y)(X'-X)] + \frac{\lambda}{2}\mathbb{E}[|X-X'|^2]. \end{aligned}$$

for $(X,Y) \sim \gamma^{opt}(\mu,\nu)$ and $X' \sim \mu'$.

Passing through super-differential and a bit of geometry

- $W_2^2(\cdot,\nu)$ is not differentiable but it is super-differentiable everywhere !
- We have the computation

$$\begin{aligned} W_2^2(\mu',\nu) - W_2^2(\mu,\nu) &\le \frac{\lambda}{2}\mathbb{E}[|X'-Y|^2] - \frac{\lambda}{2}\mathbb{E}[|X-Y|^2] \\ &= \frac{\lambda}{2}\mathbb{E}[|X'-X+X-Y|^2] - \frac{\lambda}{2}\mathbb{E}[|X-Y|^2] \\ &= \mathbb{E}[\lambda(X-Y)(X'-X)] + \frac{\lambda}{2}\mathbb{E}[|X-X'|^2]. \end{aligned}$$

for $(X,Y) \sim \gamma^{opt}(\mu,\nu)$ and $X' \sim \mu'$.

- We want to formulate $(X-Y) \in \partial^+ W_2^2(\cdot,\nu)(\mu)$.

Super-differentials in $\mathcal{P}_2(\mathbb{R}^d)$

- Keeping all the information in the previous computation, we are lead to consider elements $\psi \in \partial^+ U(\mu)$ as elements of $\{\psi : x \to \psi_x(dz) \in \mathcal{P}(\mathbb{R}^d)\}$, and for all $\mu' \in \mathcal{P}_2(\mathbb{R}^d)$ and $\Gamma \in \Pi(\mu(dx)\psi_x(dz), \mu')$

$$U(\mu') - U(\mu) \leq \int_{(\mathbb{R}^d)^3} z \cdot (x' - x)\Gamma(dx, dz, dx') + o\left(\left(\int_{(\mathbb{R}^d)^3} |x - x'|^2 \Gamma(dx, dz, dx')\right)^{\frac{1}{2}}\right).$$

Super-differentials in $\mathcal{P}_2(\mathbb{R}^d)$

- Keeping all the information in the previous computation, we are lead to consider elements $\psi \in \partial^+ U(\mu)$ as elements of $\{\psi : x \to \psi_x(dz) \in \mathcal{P}(\mathbb{R}^d)\}$, and for all $\mu' \in \mathcal{P}_2(\mathbb{R}^d)$ and $\Gamma \in \Pi(\mu(dx)\psi_x(dz), \mu')$

$$U(\mu') - U(\mu) \leq \int_{(\mathbb{R}^d)^3} z \cdot (x' - x)\Gamma(dx, dz, dx') + o\left(\left(\int_{(\mathbb{R}^d)^3} |x - x'|^2 \Gamma(dx, dz, dx')\right)^{\frac{1}{2}}\right).$$

- Trying to be more coherent with the fact that $D_\mu U(\mu) \in L^2_\mu$, we use barycentric projection and consider in fact $\phi(x) = \int_{\mathbb{R}^d} z\psi_x(dz)$ and arrive at $\phi \in \partial^+ U(\mu)$ when for all $\gamma \in \Pi(\mu, \mu')$

$$U(\mu') - U(\mu) \leq \int_{(\mathbb{R}^d)^2} \phi(x) \cdot (x' - x)\gamma(dx, dx') + o\left(\left(\int_{(\mathbb{R}^d)^2} |x - x'|^2 \gamma(dx, dz, dx')\right)^{\frac{1}{2}}\right).$$

The correct notion of viscosity solution

- Building on the previous notion of super-differential, elements we want to consider are of the form $\psi : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d)$.

The correct notion of viscosity solution

- Building on the previous notion of super-differential, elements we want to consider are of the form $\psi : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d)$.
- We have the definition

Definition (B. 23)

An upper semi-continuous function U is a viscosity sub-solution of the HJ equation if for any $\mu, \psi \in \partial^+_\mu U(\mu)$, we have

$$\lambda U(\mu) + \int_{\mathbb{R}^d \times \mathbb{R}^d} H(x, z, \mu)\psi_x(dz)\mu(dx) \leq 0.$$

We changed a bit the equation, as we are evaluating it on objects of different natures.

Theorem (B. 23)

If U and V are resp. sub and super-solution, then $U \leq V$.

Proof of the comparison principle

- Consider U sub-solution, V super-solution and $(\mu_\epsilon, \nu_\epsilon)$ a point of maximum of

$$(\mu, \nu) \to U(\mu) - V(\nu) - \frac{1}{2\epsilon} W_2^2(\mu, \nu).$$

Proof of the comparison principle

- Consider U sub-solution, V super-solution and $(\mu_\epsilon, \nu_\epsilon)$ a point of maximum of

$$(\mu, \nu) \to U(\mu) - V(\nu) - \frac{1}{2\epsilon} W_2^2(\mu, \nu).$$

- Using the definition of viscosity solution, we can estimate for γ_ϵ^o an optimal coupling between μ_ϵ and ν_ϵ

$$\begin{aligned}\lambda \sup(U - V) &\leq \lambda(U(\mu_\epsilon) - V(\nu_\epsilon)) \\ &\leq \int_{(\mathbb{R}^d)^2} H\left(x, \frac{x-y}{\epsilon}, \mu_\epsilon\right) \gamma_\epsilon^o(dx, dy) \\ &\quad - \int_{(\mathbb{R}^d)^2} H\left(y, \frac{x-y}{\epsilon}, \nu_\epsilon\right) \gamma_\epsilon^o(dx, dy)\end{aligned}$$

Proof of the comparison principle

- Consider U sub-solution, V super-solution and $(\mu_\epsilon, \nu_\epsilon)$ a point of maximum of

$$(\mu, \nu) \to U(\mu) - V(\nu) - \frac{1}{2\epsilon} W_2^2(\mu, \nu).$$

- Using the definition of viscosity solution, we can estimate for γ_ϵ^o an optimal coupling between μ_ϵ and ν_ϵ

$$\begin{aligned}
\lambda \sup(U - V) &\leq \lambda(U(\mu_\epsilon) - V(\nu_\epsilon)) \\
&\leq \int_{(\mathbb{R}^d)^2} H\left(x, \frac{x-y}{\epsilon}, \mu_\epsilon\right) \gamma_\epsilon^o(dx, dy) \\
&\quad - \int_{(\mathbb{R}^d)^2} H\left(y, \frac{x-y}{\epsilon}, \nu_\epsilon\right) \gamma_\epsilon^o(dx, dy)
\end{aligned}$$

- If $H(x, p, \mu) - H(y, p, \nu) \leq C(1 + |p|)(|x - y| + W_2(\mu, \nu))$, then we obtain $\sup(U - V) \leq C(W_2(\mu_\epsilon, \nu_\epsilon) + \frac{1}{\epsilon} W_2^2(\mu_\epsilon, \nu_\epsilon)) \to 0$

Variations on the theme

Test functions and regularization of the distance

- In several situations a much more singular term appears in the Hamiltonian

Test functions and regularization of the distance

- In several situations a much more singular term appears in the Hamiltonian
- We would then prefer to work with smooth test functions, rather than with super-differentials

Test functions and regularization of the distance

- In several situations a much more singular term appears in the Hamiltonian
- We would then prefer to work with smooth test functions, rather than with super-differentials
- This calls for the need to regularize the function

$$\mu \to W_2^2(\mu, \nu).$$

Test functions and regularization of the distance

- In several situations a much more singular term appears in the Hamiltonian
- We would then prefer to work with smooth test functions, rather than with super-differentials
- This calls for the need to regularize the function

$$\mu \to W_2^2(\mu, \nu).$$

- Main starting point, it is already semi concave !

Approximation of W_2^2

Theorem (B. and Lions '24)

For any $\nu \in \mathcal{P}_2(\Omega)$, the sequence $(\Phi_\delta)_{\delta>0}$ defined by

$$\Phi_\delta(\mu) = \sup_{\mu'} \left\{ W_2^2(\mu', \nu) - \frac{1}{\delta} W_2^2(\mu', \mu) \right\},$$

is $\mathcal{C}^{1,1}$ and converges locally uniformly toward $W_2^2(\cdot, \nu)$. If $D_\mu W_2^2(\mu, \nu)$ exists, then for any μ_δ such that $W_2^2(\mu_\delta, \mu) \to 0$, we have

$$\int_{\Omega^2} |D_\mu W_2^2(\mu,\nu)(x) - D_\mu \Phi_\delta(\mu_\delta)(x')|^2 \gamma_\delta(dx, dx') \to_{\delta\to 0} 0.$$

Main idea from Lasry-Lions regularization in Hilbert spaces.

Approximation of W_2^2

Theorem (B. and Lions '24)

For any $\nu \in \mathcal{P}_2(\Omega)$, the sequence $(\Phi_\delta)_{\delta>0}$ defined by

$$\Phi_\delta(\mu) = \sup_{\mu'} \left\{ W_2^2(\mu', \nu) - \frac{1}{\delta} W_2^2(\mu', \mu) \right\},$$

is $\mathcal{C}^{1,1}$ and converges locally uniformly toward $W_2^2(\cdot, \nu)$. If $D_\mu W_2^2(\mu, \nu)$ exists, then for any μ_δ such that $W_2^2(\mu_\delta, \mu) \to 0$, we have

$$\int_{\Omega^2} |D_\mu W_2^2(\mu, \nu)(x) - D_\mu \Phi_\delta(\mu_\delta)(x')|^2 \gamma_\delta(dx, dx') \to_{\delta \to 0} 0.$$

Main idea from Lasry-Lions regularization in Hilbert spaces.
The same optimization problem is considered in Gallouët, Natale and Todeschi (2025 and 2024), for extrapolation purposes. Somehow extrapolating is not that far from regularizing...

Elements of proof

- For the convergence of the function

$$\Phi_\delta(\mu) = \sup_{\mu'} \left\{ W_2^2(\mu', \nu) - \frac{1}{\delta} W_2^2(\mu', \mu) \right\}$$
$$= \sup_{\gamma'} \inf_{\gamma} \left\{ \int_{(\mathbb{R}^d)^2} |x-y|^2 \gamma'(dx, dy) - \frac{1}{\delta} \int_{(\mathbb{R}^d)^2} |x-y|^2 \gamma(dx, dy) \right\}$$
$$\leq \inf_{\gamma} \sup_{\gamma'} \cdots = \frac{1}{1-\delta} W_2^2(\mu, \nu).$$

- The convergence of the gradient follows from the fact that $W_2^2(\cdot, \nu)$ is semi-concave. As in standard Hilbert space, this implies that its super-differential is closed, from which follows the stability of the derivative.

The regularity of Φ_δ

- We want to state that $\Phi_\delta \in \mathcal{C}^{1,1}$. This implies that we need to compare $D_\mu \Phi_\delta(\mu) \in L^2_\mu$ and $D_\mu \Phi_\delta(\nu) \in L^2_\nu$ for $\nu \neq \mu$.

The regularity of Φ_δ

- We want to state that $\Phi_\delta \in \mathcal{C}^{1,1}$. This implies that we need to compare $D_\mu \Phi_\delta(\mu) \in L^2_\mu$ and $D_\mu \Phi_\delta(\nu) \in L^2_\nu$ for $\nu \neq \mu$.
- For that, a suitable notion of parallel transport is needed.

The regularity of Φ_δ

- We want to state that $\Phi_\delta \in \mathcal{C}^{1,1}$. This implies that we need to compare $D_\mu \Phi_\delta(\mu) \in L^2_\mu$ and $D_\mu \Phi_\delta(\nu) \in L^2_\nu$ for $\nu \neq \mu$.
- For that, a suitable notion of parallel transport is needed.
- Previous work from Ambrosio and Gigli brings partial answer.

The regularity of Φ_δ

- We want to state that $\Phi_\delta \in \mathcal{C}^{1,1}$. This implies that we need to compare $D_\mu \Phi_\delta(\mu) \in L^2_\mu$ and $D_\mu \Phi_\delta(\nu) \in L^2_\nu$ for $\nu \neq \mu$.
- For that, a suitable notion of parallel transport is needed.
- Previous work from Ambrosio and Gigli brings partial answer.

Statement of the problem

Given μ, ν and a curve θ between them, we want to (parallel) transport $\psi \in L^2_\mu$ along θ so that we can compare it with elements of L^2_ν.

Parallel transport

Theorem (B. 25)

Given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, a continuous (Lagrangian) curve θ in $\mathcal{P}_2(\mathbb{R}^d)$ and $\psi \in L^2_\mu$, there exists a well defined parallel transport of ψ along θ.

- *It is a Lagrangian curve $\Psi \in \mathcal{P}([0,1] \to \mathbb{R}^d \times \mathbb{R}^d)$.*
- $(e_0)_\# \Psi = (Id, \psi)_\# \mu$.
- $(\pi_2)_\# \Psi$ *is concentrated on constante curves.*
- $(e_1)_\# \Psi$ *is not necessary of the form $(Id, \tilde{\psi})_\# \nu$, but simply satisfies that its first marginal is ν.*

The main difference with the existing literature is that it is a Lagrangian point of view, quite more powerful.

A precise definition of $\mathcal{C}^{1,1}$

- With this notion of parallel transport, we can indeed compare derivatives at different points.

A precise definition of $\mathcal{C}^{1,1}$

- With this notion of parallel transport, we can indeed compare derivatives at different points.
- We parallel transport $\phi \in L^2_\mu$ toward ν, along a coupling $\gamma \in \Pi(\mu, \nu)$, and we evaluate the difference at ν.

A precise definition of $\mathcal{C}^{1,1}$

- With this notion of parallel transport, we can indeed compare derivatives at different points.
- We parallel transport $\phi \in L^2_\mu$ toward ν, along a coupling $\gamma \in \Pi(\mu,\nu)$, and we evaluate the difference at ν.
- This yields

$$\int_{(\mathbb{R}^d)^2} |D_\mu U(\mu)(x) - D_\mu U(\nu)(y)|^2 \gamma(dx,dy) \leq$$
$$\leq C \int_{(\mathbb{R}^d)^2} |x-y|^2 \gamma(dx,dy).$$

A precise definition of $\mathcal{C}^{1,1}$

- With this notion of parallel transport, we can indeed compare derivatives at different points.
- We parallel transport $\phi \in L^2_\mu$ toward ν, along a coupling $\gamma \in \Pi(\mu,\nu)$, and we evaluate the difference at ν.
- This yields

$$\int_{(\mathbb{R}^d)^2} |D_\mu U(\mu)(x) - D_\mu U(\nu)(y)|^2 \gamma(dx,dy) \leq$$
$$\leq C \int_{(\mathbb{R}^d)^2} |x-y|^2 \gamma(dx,dy).$$

- Equivalent to $C^{1,1}$ regularity of the lift.

Creating points of maximum

- I jumped over the fact that $\mathcal{P}_2(\mathbb{R}^d)$ is not locally compact...

Creating points of maximum

- I jumped over the fact that $\mathcal{P}_2(\mathbb{R}^d)$ is not locally compact...
- Fundamental in the proof of comparison to be able to consider points of maximum of

$$\mu \to U(\mu) - \lambda W_2^2(\mu, \nu),$$

for $\mu \to U(\mu)$ which is simply usc.

Creating points of maximum

- I jumped over the fact that $\mathcal{P}_2(\mathbb{R}^d)$ is not locally compact...
- Fundamental in the proof of comparison to be able to consider points of maximum of

$$\mu \to U(\mu) - \lambda W_2^2(\mu, \nu),$$

for $\mu \to U(\mu)$ which is simply usc.
- We can assume quite easily that U is bounded, thus we know that we can restrict our search on $B(\nu, \lambda^{-1}\|U\|_\infty)$, which is bounded but not compact...

A variant of Stegall's Lemma

We need to consider the following function

$$\mathcal{I}(\mu,\nu) := - \sup_{\gamma\in\Pi(\mu,\nu)} \int_{(\mathbb{R}^d)^2} x\cdot y \quad \gamma(dx,dy).$$

Proposition (B. - Lions 25)

For any bounded usc $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, $\lambda > 0$ *and* $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, *for any* $\epsilon > 0$, *there exists* $\nu_\epsilon \in \mathcal{P}_2(\mathbb{R}^d)$ *such that* $\int_{\mathbb{R}^d} |x|^2 \nu_\epsilon(dx) < \epsilon$ *and*

$$\mu \to U(\mu) - \lambda W_2^2(\mu,\nu) + \mathcal{I}(\mu,\nu_\epsilon)$$

has a unique point of exposed maximum in $\mathcal{P}_2(\mathbb{R}^d)$ *at some* μ_ϵ.

Proof of Stegall's Lemma

- Lift everything to $L^2(\Omega, \mathbb{R}^d)$: $\mathcal{U}(X) = U(\mathcal{L}(X))$.
- Stegall's Lemma is valid in Hilbert spaces.

Proof of Stegall's Lemma

- Lift everything to $L^2(\Omega, \mathbb{R}^d) : \mathcal{U}(X) = U(\mathcal{L}(X))$.
- Stegall's Lemma is valid in Hilbert spaces.
- Consider Y_ϵ, such that $\|Y_\epsilon\| \leq \epsilon$ and

$$X \to \mathcal{U}(X) + \mathbb{E}[Y \cdot X]$$

has a unique point of exposed maximum.

Proof of Stegall's Lemma

- Lift everything to $L^2(\Omega, \mathbb{R}^d) : \mathcal{U}(X) = U(\mathcal{L}(X))$.
- Stegall's Lemma is valid in Hilbert spaces.
- Consider Y_ϵ, such that $\|Y_\epsilon\| \leq \epsilon$ and

$$X \to \mathcal{U}(X) + \mathbb{E}[Y \cdot X]$$

  has a unique point of exposed maximum.
- Pulling anything down we end up with

$$\mu \to U(\mu) + \mathcal{I}(\mu, \mathcal{L}(-Y_\epsilon))$$

  has a unique point of strict maximum.

Equations on $\mathcal{M}_+(\Omega)$

- Consider the case where we want to control

$$\partial_t m + \mathrm{div}(\alpha m) + \lambda m = 0$$

with a cost

$$\inf_{\alpha,\lambda,m} \int_0^T \int_\Omega \{L_1(x,\alpha_t(x)) + L_2(x,\lambda_t(x))\} m_t(dx) + G(m_T).$$

- HJB is given by

$$-\partial_t U + \int_\Omega H_1(x, D_\mu U) + H_2(x, \nabla_\mu U)\mu(dx) = 0 \text{ in } (0,T)\times\mathcal{M}_+(\Omega).$$

- The use of W_2^2 is not possible anymore...
- With G. Ceccherini, we replace W_2^2 with Hellinger-Kantorovich distances or Wasserstein-Fisher-Rao. This work is almost done!

Equations on $\mathcal{M}_+(\Omega)$ II

- With G. Ceccherini, we replace W_2^2 with Hellinger-Kantorovich or Wasserstein-Fisher-Rao distances.
- Somehow, in the doubling of variables, we are interested in

$$(\mu, \nu) \mapsto U(\mu) - V(\nu) - \mathcal{A}(\epsilon, \mu, \nu)$$

-

Equations on $\mathcal{M}_+(\Omega)$ II

- With G. Ceccherini, we replace W_2^2 with Hellinger-Kantorovich or Wasserstein-Fisher-Rao distances.
- Somehow, in the doubling of variables, we are interested in

$$(\mu, \nu) \mapsto U(\mu) - V(\nu) - \frac{1}{2\epsilon} W_2^2(\mu, \nu)$$

-

Equations on $\mathcal{M}_+(\Omega)$ II

- With G. Ceccherini, we replace W_2^2 with Hellinger-Kantorovich or Wasserstein-Fisher-Rao distances.
- Somehow, in the doubling of variables, we are interested in

$$(\mu, \nu) \mapsto U(\mu) - V(\nu) - \frac{1}{2\epsilon} WFR_2^2(\mu, \nu)$$

- We made this heuristic precise in smooth cases.
- Properties of general control problems follows as usual from **Geometric** properties of the state space

Bibliography

- Similar HJB equations on the Wasserstein space : Gangbo, Tudorascu, Mayorga, Swiech, Cardaliaguet, Quincampoix, Jimenez, Marigonda, ...
- With Fokker-Planck equations : Feng, Soner et al., Cosso, Gozzi, Kharroubi, Pham, Cecchin and Delarue, Zhang, Conforti, Kraaij, Tonon, Daudin, Seeger, Jackson...
- Optimality conditions : Benamou, Brenier, Daudin, Bonnet, Frankowska, Russo...
- Differential calculus, regularity : Ambrosio-Gigli-Savaré, Lions, Otto, Gangbo, Tudorascu, Alfonsi, Jourdain...

A few works on the topic

- Stochastic optimal transport and HJB equations on the set of probability measures, CB, 2024, to appear in AIHPC.
- CB, PL Lions, An approximation of the squared Wasserstein distance and an application to HJ equation, 2024, arxiv.
- CB, PL Lions and P.E. Souganidis Optimal control of the Dyson equation and large deviations of random matrices, forthcoming.
- CB and G Ceccherini, HJB equations on the set of positive measures, forthcoming.
- CB, Little book/lecture notes on PDE on the space of probability measures, forthcoming.

Thank you for your the invitation

Thank you for your attention