Hamilton-Jacobi PDE in space of Probability Measures

Jin Feng

Department of Mathematics, University of Kansas, U.S.A.

June 5th, Kac Lecture at Utrecht University, The Netherlands

Outline

1. Prelude
   - Freidlin-Wentzell theory for singular diffusions
   - Space of probability measures as a quotient space
2. Variational problems in space of probability measures
   - Hamilton-Jacobi equations in $\mathcal{P}_2(\mathbb{R}^d)$
   - Why are they interesting?
3. The LDP Problem (A simplified version):
   - Optimal transport
4. The Mechanics problem
   - H-J equation in $\mathcal{P}_2(\mathbb{R}^d)$ - inadequate choice of tangent space
   - H-J equation in $\mathcal{P}_2(\mathbb{R}^d)$ – Geometric tangent cone

Small diffusions

- $dX_n(t) = b(X_n(s))ds + \frac{1}{\sqrt{n}}\sigma(X_n(s))dW(s)$.
- $A_n f(x) = \frac{1}{2n}\sum_{ij} a_{ij}(x)\partial^2_{ij} f(x) + b(x)\cdot\nabla f(x)$
- $H_n f(x) = \frac{1}{2n}\sum_{ij} a_{ij}(x)\partial^2_{ij} f(x) + \frac{1}{2}|\sigma(x)\nabla f(x)|^2 + b(x)\cdot\nabla f(x)$,
- $H_n f \to Hf$ with

$$Hf(x) = \frac{1}{2}|\sigma(x)\nabla f(x)|^2 + b(x)\cdot\nabla f(x) = H(x, \nabla f(x)).$$

- $L(x,q) = \sup_p\{pq - H(x,p)\}$.
- $\{X_n, n = 1, 2, \ldots\}$ satisfies LDP with rate function

$$I(x(\cdot)) = \int_0^\infty L(x(s), \dot{x}(s))ds.$$

Comparison Principle

$$f - \alpha Hf = h.$$

By available PDE theory, the comparison principle holds if,

- σ, b are Lipschitz, OR
- σ, b are bounded continuous and $\sigma(x)\sigma^T(x)$ is uniform nondegenerate.

Comparison Principle – singular case

$$f - \alpha Hf = h.$$

where $E = \mathbb{R}_+$ and $\sigma(x) = \sqrt{x}$ and

$$Hf(x) = \frac{1}{2} x |\partial_x f|^2.$$

- The comparison principle is NOT known.
- Small time heat kernel estimate for $dX(t) = b(X(t))dt + \sigma(X(t))dW(t)$:

$$\lim_{\epsilon \to 0+} \lim_{n \to \infty} t \log P(X(n^{-1}t) \in B(y, \epsilon) | X(0) = x) = -\frac{d^2(x,y)}{t}.$$

- Rescaling: let $X_n(t) = X(n^{-1}t)$.

Comparison principle does hold for the singular case

- A geometrically invariant condition:

$$H_x \frac{d^2(x,y)}{\epsilon} - H_y \frac{-d^2(x,y)}{\epsilon} \leq \omega\left(\frac{d^2(x,y)}{\epsilon} + d^2(x,y)\right).$$

- Take $E := (\mathbb{R}_+, d)$ where

$$d^2(x,y) := \inf\{\int_0^1 L(\dot{x}(s))ds : x(0)=x, x(1)=y\} = |\sqrt{x} - \sqrt{y}|^2.$$

- A geometric identity

$$|\sqrt{x}\nabla_x d^2(x,y)|^2 = |\sqrt{y}\nabla_y\Big(-d^2(x,y)\Big)|^2.$$

- $x \mapsto \mathrm{d}^2(x,y)$ not smooth in classical sense.
- A deeper question: Right choice of test functions?
- Need for "viscosity extensions".

Space of empirical measures:

- $x_i \in \mathbb{R}^d$, $i = 1, 2, \ldots, n$.
- Ordered $\vec{x} := (x_1, \ldots, x_n)$ v.s un-ordered $\{x_1, \ldots, x_n\}$.
- Un-ordered n-points as a permutation invariant element in $(\mathbb{R}^d)^n / \sim$:

$$(x_1, \ldots, x_n) \sim (x_{\pi(1)}, \ldots x_{\pi(n)}).$$

- Representing $(\mathbb{R}^d)^n / \sim$ as

$$E_n \quad := \quad \{\rho(dx) := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}(dx), x_i \in \mathbb{R}^d\}.$$

- Let $E := \mathcal{P}(\mathbb{R}^d)$ be a kind of limit of E_ns.
- Large deviation, limits and dynamics of HJ equations in E_n and in E.
- Very singular spaces (many "corners" and "edge") of metric geometry nature.
- Metric space analysis tools is natural – theory of mass transport.

Outline

1. Prelude
   - Freidlin-Wentzell theory for singular diffusions
   - Space of probability measures as a quotient space
2. Variational problems in space of probability measures
   - Hamilton-Jacobi equations in $\mathcal{P}_2(\mathbb{R}^d)$
   - Why are they interesting?
3. The LDP Problem (A simplified version):
   - Optimal transport
4. The Mechanics problem
   - H-J equation in $\mathcal{P}_2(\mathbb{R}^d)$ - inadequate choice of tangent space
   - H-J equation in $\mathcal{P}_2(\mathbb{R}^d)$ – Geometric tangent cone

Lagrangian in space of probability measures

- $(\mathsf{X}, \mathsf{d}) := (\mathcal{P}_2(\mathbb{R}^d); W_2)$ Wasserstein order-2 metric space,

$$A_T[\rho(\cdot)] := \int_0^T L(\rho, \dot\rho) dt.$$

- A Large Deviation Problem: $S(\rho) := \int \rho \log \rho + \int \Psi d\rho - \log Z_\Psi$ and

$$L(\rho, \dot\rho) := \frac{1}{2} \|\dot\rho + \mathrm{grad} S(\rho)\|^2_{-1,\rho}.$$

- A Mechanics Problem: ϕ, Φ smooth and bounded and

$$L(\rho, \dot\rho) := \frac{1}{2} \|\dot\rho\|^2_{-1,\rho} - V(\rho), \quad V(\rho) := \int_{\mathbb{R}^d} \phi d\rho + \frac{1}{2}\langle \Phi * \rho, \rho \rangle.$$

Hamiltonian

All Hamiltonians are discontinuous and have a strong form of singularity

- The Large Deviation Problem:

$$H(\rho, \mathrm{grad}_\rho f) = \langle -\mathrm{grad} S, \mathrm{grad} f \rangle_{-1,\rho} + \frac{1}{2}\|\mathrm{grad} f\|^2_{-1,\rho}.$$

Feature: Controlled gradient flow in $\mathcal{P}_2(\mathbb{R}^d)$.

- The Mechanics Problem

$$H(\rho, \mathrm{grad}_\rho f) = \frac{1}{2}\|\mathrm{grad} f\|^2_{-1,\rho} + V(\rho).$$

Feature: Condensation.

Hamilton-Jacobi equations

In the talk, we solve well-posedness for $f - \alpha H(\rho, \mathrm{grad}_\rho f) = h$.

- The Large Deviation Problem: F. &Katsoulakis [ARMA 09], F. & Kurtz [AMS Book 06], F. & Nguyen [JMPA12].
- The Mechanics Problem: Ambrosio & F. [JDE 14]. Earlier: Hynd, Kim, Gangbo, Nguyen, Swiech, Tudorascu

Why are they interesting?

The Mechanics Problem:

- Variational formulation of compressible Euler equation

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho u) & = & 0 \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u) & = & -\rho \nabla(\phi + \Phi * \rho). \end{cases}$$

- Formally, Hamiltonian flow as time-dependent gradient flow:

$$\dot{\rho} \in \mathrm{grad}_\rho S(t, \rho)$$

- Understand certain behaviors of compressible Euler equation in $\mathbb{R}^d$ by lifting it up to the level of Hamilton-Jacobi equation in $\mathcal{P}_2(\mathbb{R}^d)$.

Why are they interesting?

The Large Deviation Problem:

- Mean-field interacting diffusions (more general than the earlier model)

$$dX_i = K * \rho_n(t, X_i)dt + \sqrt{2\nu}dW_i(t), \ \rho_n(t) := \frac{1}{n}\sum_i \delta_{X_i(t)}.$$

- LDP rate function as Boltzmann entropy in path space

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} -\frac{1}{n}\log P(\rho_n(\cdot) \in B_\epsilon(\rho(\cdot)) = A_T(\rho(\cdot)).$$

- Feng&Kurtz [06] "Large Deviation for Stochastic Processes" :
  1. Hamiltonian for Markov process $\{\rho_n(\cdot)\}$: $H_n f(\rho) := \frac{1}{n}e^{-nf}A_n e^{nf}$;
  2. $H_n \to H$;
  3. Exp. tightness;
  4. Comparison principle for resolvent eqn of H.

Outline

LDP for infinite particles

- LDP for $dX_i = \sqrt{2} dW_i - \nabla\Psi(X_i)dt$,

$$\rho_N(t, dx) = N^{-1} \sum_i \delta_{X_i(t)}(dx).$$

- For "polynomial" fs:

$$Hf(\rho) = \langle \Delta\rho + \mathrm{div}(\rho\nabla\Psi), \frac{\delta f}{\delta\rho} \rangle + \frac{1}{2} \int |\nabla \frac{\delta f}{\delta\rho}|^2 d\rho.$$

- The above is in fact,

$$Hf(\rho) = \langle -\mathrm{grad} S_\Psi(\rho), \mathrm{grad} f(\rho) \rangle_{-1,\rho} + \frac{1}{2} \|\mathrm{grad} f(\rho)\|^2_{-1,\rho}.$$

- What is the supporting notion of differential structure ?
- The answer: The Otto calculus, formalized by Ambrosio-Gigli-Savare.

The tangent space of $\mathcal{P}_2(\mathbb{R}^d)$

$E := \mathcal{P}_2(\mathbb{R}^d)$.

- Identify $v \in T_\rho E$ in $\partial_t \rho + \mathrm{div}(\rho v) = 0$ with $\dot\rho = \partial_t \rho$. How?
- $\|m\|^2_{-1,\rho} := \sup_{\varphi \in C_c^\infty(\mathbb{R}^d)} \{2\langle \varphi, m\rangle - \int_{\mathbb{R}^d} |\nabla \varphi|^2 d\rho\}$,
- $\dot\rho \in H_{-1,\rho}(\mathbb{R}^d) := \{m \in \mathcal{D}'(\mathbb{R}^d) : \|m\|_{-1,\rho} < \infty\}$

$$\Leftrightarrow v \in T_\rho E.$$

A differential structure on $\mathcal{P}_2(\mathbb{R}^d)$

- Directional derivative along smooth direction $p \in C_c^\infty$

$$D^p f(\rho(0)) := \lim_{t \to 0+} \frac{f(\rho(t)) - f(\rho(0))}{t},$$

$$\partial_t \rho + \operatorname{div}(\rho \nabla p) = 0.$$

- Gradient

$$D^p f(\rho(0)) = \langle \text{grad } f(\rho(0)), p \rangle.$$

- Why? A property

$$\langle m, -\nabla(\rho \nabla p) \rangle_{-1,\rho} = \langle m, p \rangle.$$

-

$$\text{grad } f(\rho) = -\nabla(\rho \nabla \frac{\delta f}{\delta \rho}).$$

The Otto calculus on $\mathcal{P}_2(\mathbb{R}^d)$

- Let $S(\rho) = \int \rho \log \rho dx$, then
$$-\Delta \rho = -\nabla \cdot (\rho \frac{\nabla \rho}{\rho}) = \text{grad } S.$$
- Let
$$S_\Psi(\rho) = S(\rho) + \int \Psi d\rho + \log Z = \int \log \frac{d\rho}{d\mu^\Psi} d\rho,$$
$\mu^\Psi = Z^{-1} e^{-\Psi} dx$. Then
$$\text{grad } S_\Psi(\rho) = -\Delta \rho - \nabla(\rho \nabla \Psi).$$
- Fokker-Planck equation is $\dot{\rho} = -\text{grad} S_\Psi$.

Comparison principle holds

$$(I - \alpha H)f = h$$

- $Hf(\rho) = \langle -\text{grad } S_\Psi, \text{grad } f\rangle_{-1,\rho} + \frac{1}{2}\|\text{grad} f(\rho)\|^2_{-1,\rho}$.
- Optimal controlled gradient flow

$$\dot{\rho} = -\text{grad } S_\Psi(\rho) + m, \quad L(\rho, m) = \frac{1}{2}\|m\|^2_{-1,\rho}.$$

- F.&Kurtz and F.&Katsoulakis : This HJ equation is well-posed, in particular, the comparison holds.

Choice of test functions, useful for comparison

- Choice of test functions:

$$f_0(\rho) = \alpha d^2(\rho,\gamma) + \epsilon S(\rho),$$

$$f_1(\gamma) = -\alpha d^2(\rho,\gamma) - \epsilon S(\gamma).$$

- Lyapunov functional: (I is Fisher information)

$$H\epsilon S(\rho) = -\epsilon(1 - \frac{1}{2}\epsilon) I(\rho).$$

- Wasserstein metric
  $d^2(\rho,\gamma) := \inf\{\int_0^1 \|\dot{\rho}\|^2_{-1,\rho(r)} dr : \rho(0) = \rho, \rho(1) = \gamma\}$.
- Brenier, Otto, Ambrosio-Gigli-Savare, Villani....

$$\|\text{grad}_\rho \frac{1}{2} d^2(\rho,\gamma)\|^2_{-1,\rho} = d^2(\rho,\gamma) = \|\text{grad}_\gamma \frac{1}{2} d^2(\rho,\gamma)\|^2_{-1,\gamma}.$$

Estimate

- HWI inequality:

$$S_\Psi(\rho) - S_\Psi(\gamma) \leq C_\Psi(d(\rho,\gamma)\sqrt{I(\gamma)} + d^2(\rho,\gamma)).$$

- Fisher information

$$I(\rho) = \|\text{grad } S_\Psi\|^2_{-1,\rho} = \int \frac{|\nabla\rho|^2}{\rho}dx + \int (|\nabla\Psi|^2 - 2\Delta\Psi)d\rho.$$

- The strong regularization of I helps (mass transport HWI inequalities...)

$$Hf_0(\rho) - Hf_1(\gamma) \leq \omega(\alpha d^2(\rho,\gamma)).$$

Explained using Euclidean space analogy:

$$Hf(x) := -\nabla S(x)\nabla f(x) + \frac{1}{2}|\nabla f(x)|^2.$$

- Mass transport version of: $|\nabla_x |x-y|^2| = |\nabla_y(-|x-y|^2)|$.
- Contraction of flows by $-\nabla S$, with respect to $d(x,y) := |x-y|$

$$-\nabla S(x)\nabla_x d^2(x,y) - \nabla S(y)\nabla_y d^2(x,y) \leq C\mathrm{d}^2(x,y).$$

The case of $E := \mathcal{P}(\mathbb{R}^d)$

- If ρ has Lebesgue density, by Brenier's theory on mass transport

$$\mathrm{grad}_\rho \frac{\mathrm{d}^2(\rho,\gamma)}{2} = \nabla p, \quad p(x) := \frac{|x|^2}{2} - \varphi(x),$$

where $\varphi := \varphi^{\rho,\gamma}$ is the convex optimal potential in the Monge formulation of d such that $(\nabla\varphi)_\# \rho = \gamma$. In the above,

$$\|\mathrm{grad}_\rho \frac{\mathrm{d}^2(\rho,\gamma)}{2}\|_\rho^2 = \int |\nabla p|^2 d\rho = \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2 \rho(dx) = \mathrm{d}^2(\rho,\gamma).$$

- Log-Sobolev type/HWI inequality

$$\langle \mathrm{grad}_\rho \mathrm{d}^2, \mathrm{grad}_\rho(-S)\rangle_\rho + \langle \mathrm{grad}_\gamma \mathrm{d}^2, \mathrm{grad}_\gamma(-S)\rangle_\gamma \le C\mathrm{d}^2$$

- The strong regularization of I helps to give

$$Hf_0(\rho) - Hf_1(\gamma) \le \omega(\alpha \mathrm{d}^2(\rho,\gamma) + \mathrm{d}(\rho,\gamma)).$$

Outline

1. Prelude
   - Freidlin-Wentzell theory for singular diffusions
   - Space of probability measures as a quotient space
2. Variational problems in space of probability measures
   - Hamilton-Jacobi equations in $\mathcal{P}_2(\mathbb{R}^d)$
   - Why are they interesting?
3. The LDP Problem (A simplified version):
   - Optimal transport
4. The Mechanics problem
   - H-J equation in $\mathcal{P}_2(\mathbb{R}^d)$ - inadequate choice of tangent space
   - H-J equation in $\mathcal{P}_2(\mathbb{R}^d)$ – Geometric tangent cone

H-J equation for compressible Euler equations

The continuum mechanics problem

$$Hf(\rho) := \frac{1}{2}\|\text{grad} f(\rho)\|^2_{-1,\rho} + V(\rho).$$

- Earlier attempts by Gangbo, Nguyen, Tadurascu, Swiech, Hydn, Kim, ... on equivalent re-formulation

$$L(\rho, v) := \frac{1}{2}\int_{\mathcal{O}} |v|^2 d\rho - V(\rho), \quad H(x, \xi) := \frac{1}{2}\int_{\mathcal{O}} |\xi|^2 d\rho + V(\rho)$$

  where $\mathcal{O} := \mathbb{T}^d, \mathbb{R}^1, \mathbb{R}^d$.
- Using sub- super-gradients in Wasserstein space.
- Open Problem: No uniqueness theory (i.e. comparison principle).

It should not be expected to work, metric nature of $\mathcal{P}_2(\mathbb{R}^d)$

- Kantorovich formulation of

$$\mathrm{d}^2(\rho,\gamma) := \inf_{\pi^1_\# m=\rho, \pi^2_\# m=\gamma} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 m(dx,dy)$$

  $\Gamma(\rho,\gamma) :=$ those optimal m in the inf problem.

- For every $m \in \Gamma(\rho,\gamma)$, let

$$m(dx,dy) := m(dy|x)\rho(dx), \quad u(x) := \int_{\mathbb{R}^d} (x-y) m(dy|x).$$

  Then $\mathrm{grad}_\rho \frac{\mathrm{d}^2(\rho,\gamma)}{2} = u$ (equivalent class). (By A.G.S.05)

- By Jensen, equality holds if and only if $\gamma = T_\# \rho$ (map v.s. plan)

$$\|\mathrm{grad}_\rho \frac{\mathrm{d}^2(\rho,\gamma)}{2}\|^2_\rho = \int_{\mathbb{R}^d} |u|^2 d\rho \leq \int\int |x-y|^2 m(dy|x)\rho(dx) = \mathrm{d}^2(\rho,\gamma)$$

What happened

- The T_ρ did not generate enough velocity fields.
- There is a metric analysis (hence taking care of singularity of space) based re-formulation by Ambrosio- Feng 15.
- When ρ is singular (i.e. "corned"), certain "direction" in the metric formulation cannot be modelled by the tangent space here – think of a Polyhedron.

Geometric tangent cone

Idea: relaxed formulation of tangent $\mathbf{m}(dx, d\xi) := \delta_{u(x)}(d\xi)\rho(dx)$.

1. $G(\rho) := \{\mathbf{m} \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) : (\pi^1, \pi^1 + \epsilon\pi^2)_\# \mathbf{m} \in \Gamma_o(\rho, \gamma), \exists \gamma \in X, \epsilon > 0\}$
2. $D_\rho(\mathbf{m}_1, \mathbf{m}_2) := \inf\{\int_{(\mathbb{R}^d)^3} |\xi - \eta|^2 M(dx; d\xi, d\eta) : M \in \mathcal{P}_2((\mathbb{R}^d)^3), \pi^{1,2}_\# M = \mathbf{m}_1, \pi^{1,3}_\# M = \mathbf{m}_2\}$
3. $\langle m_1, m_2 \rangle_\rho := \max\{\int_{(\mathbb{R}^d)^3} \xi\eta M(dx; d\xi, d\eta) : M \in \mathcal{P}_2((\mathbb{R}^d)^3), \pi^{1,2}_\# M = \mathbf{m}_1, \pi^{1,3}_\# M = \mathbf{m}_2\}$
4. $\|\mathbf{m}\|^2_\rho := \langle \mathbf{m}, \mathbf{m} \rangle_\rho$

Definition

$$\mathrm{Tan}_\rho := \overline{G(\rho)}^{D_\rho(\cdot,\cdot)}, \quad \mathrm{Tan} := \cup_\rho \mathrm{Tan}_\rho.$$

Geometric tangent cone - continued

Lemma

$T_\rho \hookrightarrow \mathrm{Tan}_\rho$. *When $\rho(dx) = \rho(x)dx$, the embedding is isometric and $1-1$, generally the inclusion is strict.*

sub-, super-differential calculus

1. Fréchet super-differentials for $f : X \mapsto \bar{\mathbb{R}}$. $\mathbf{n} \in \partial^+_\rho f$ if there exists a modulus $\omega_{\mathbf{n}}$ such that, for every $\rho_1 \in X$ and every $\mathbf{M} \in \mathcal{P}_2((\mathbb{R}^d)^3)$ with $(\pi^1, \pi^1 + \pi^2)_\# \mathbf{M} \in \Gamma_o(\rho, \rho_1)$ and $\pi^{1,3}_\# = n$, we have

$$f(\rho_1) - f(\rho) \le \int_{(\mathbb{R}^d)^3} (\xi \cdot \eta) \mathbf{M}(dx; d\xi, d\eta) + \mathrm{d}(\rho, \rho_1) \omega_{\mathbf{n}}(\mathrm{d}(\rho, \rho_1)) \quad (1)$$

2. Similarly define (sub-) derivative. $\mathbf{n} \in \partial^-_\rho f$.
3. $\partial f := \partial^+ f \cap \partial^- f$.
4. $\mathbf{n} \oplus \mathbf{m} := \{\mathbf{n} := (\pi^1, \pi^2 + \pi^3)_\# \mathbf{N}, \pi^{1,2}_\# \mathbf{N} = \mathbf{n}, \pi^{1,3}_\# \mathbf{N} = \mathbf{m}\}$.

Lemma

$$\partial^i \partial \varphi_1 \oplus \partial^i \varphi_2 \subset \partial^i (\varphi_1 + \varphi_2)$$

Geometric Hamiltonian

1. $f - \mathbf{H}_0 f \le h$ and $f - \mathbf{H}_1 f \ge h$.

Lagrangian and Hamiltonian – geometric tangent cone

$\mathcal{P}_2(\mathbb{R}^d)$ is a quotient space.

1. $\mathbf{m}(dx, dv) := \mathbf{m}(dv|x)\rho(dx)$ in stead of $(\rho(dx), v(x))$.
2. $\mathbf{L}(\mathbf{m}) := \frac{1}{2}\|\mathbf{m}\|_\rho^2 - V(\rho)$ and $\mathbf{H}(\mathbf{n}) := \frac{1}{2}\|\mathbf{n}\|_\rho^2 + V(\rho)$;
3. $\mathbf{H}_0 f(\rho) := \inf\{\mathbf{H}(\mathbf{n}) : \mathbf{n} \in \partial_\rho^+ f \cap \mathrm{Tan}_\rho\}$;
4. $\mathbf{H}_1 f(\gamma) := \sup\{\mathbf{H}(\mathbf{n}) : \mathbf{n} \in \partial_\gamma^- f \cap \mathrm{Tan}_\gamma\}$.

Theorem

(Ambrosio-Feng15.) The HJ equation in the geometric tangent cone formulation is well posed. In particular, comparison principle holds.