Intro :

• Def : [Gradient Flow in Linear Space]

$\overline{X}$ is a linear space, and $F : \overline{X} \to \mathbb{R}$ is smooth. Gradient flow (or steepest descent curve) is a smooth curve $x : \mathbb{R} \longrightarrow \overline{X}$ s.t. $x'(t) = -\nabla F(x(t))$.

• e.g. for $\overline{X} = L^2(\mathbb{R}^n)$, a Hilbert space, and for Dirichlet energy $F(u) = \frac{1}{2}\int |\nabla u(x)|^2 dx$, the Heat Equation $\partial_t u = \nabla^2 u$ is a gradient flow problem.

• 2. Numerical methods and their convergence : Since gradient flow gradually minimizes $F(x)$, so many optimization methods are related to it. e.g. gradient descent, proximal descent, mirror descent.

• 3. Generalization to the gradient flow to general metric space.

The need of viewing PDEs as gradient flows on general metric spaces.

• e.g. : PDEs in the continuity equation from $\partial_t \rho - \nabla\cdot(\rho v) = 0$, where $v = \nabla\left[\frac{\delta F}{\delta \rho}\right]$, can be cast as a gradient flow on the space of probability with Wasserstein distance.

The need of minimizing functionals on metric space.

• Optimization w.r.t. probability Dist e.g. $\min_q KL(q \| p)$. Optimization without parameterization is possible.

(Stein Variational Gradient Descent).

Gradient Flow in Euclidean Space.

Variants of Gradient Flow in the Euclidean Space.

- Variants 0: $F : \mathbb{R}^n \to \mathbb{R}$ is differentiable (Cauchy pbm):

$$\begin{cases} x'(t) = -\nabla F(x(t)), \text{ for } t>0 \\ x(0) = x_0 \end{cases}$$

- Thm: exists unique solution if ∇F is Lipschitz.

- Variants 1: F is convex and unnecessarily differentiable:

$$\begin{cases} x'(t) \in -\partial F(x(t)) \quad \text{for a.e. } t>0, \\ x(0) = x_0. \end{cases}$$

where x is an absolutely continuous curve, and $\partial F(x) = \{ p \in \mathbb{R}^n : \forall y \in \mathbb{R}^n, F(y) \geq F(x) + p\cdot(y-x) \}$.

- Thm: Any two solutions x_1, x_2 of the above problem with different initial conditions

satisfy $|x_1(t) - x_2(t)| \leq |x_1(0) - x_2(0)|$

- Cor: For a given initial condition, the above problem has one unique solution.

- Variant 2: F is semi-convex (λ-convex).

- Def [λ-convex function]: F is λ-convex ($\lambda \in \mathbb{R}$) if $F(x) - \frac{\lambda}{2}|x|^2$

$$\begin{cases} x'(t) \in -\partial F(x(t)), \text{ for a.e. } t>0 \\ x(0) = x_0. \end{cases}$$

where x is an absolutely continuous curve, and

$\partial F(x) = \{ p \in \mathbb{R}^n : \forall y \in \mathbb{R}^n, F(y) \geq F(x) + p\cdot(y-x) + \frac{\lambda}{2}|y-x|^2 \}$.

• Thm : Any two solutions x_1, x_2 of the above problem with different initial conditions satisfy

$$|x_1(t) - x_2(t)| \le e^{-\lambda t} |x_1(0) - x_2(0)|.$$

• Cor : • For a given initial condition, the above pbm has one unique solution.

• If $\lambda > 0$ (strong convex), F has a unique minimizer x^*. $x(t) \equiv x^*$ is a solution, so for any solution $x(t)$, $|x(t) - x^*| \le e^{-\lambda t} |x(0) - x^*|$.

Approximating Curves

• Def [MMS]: Minimizing Movement Scheme (MMS): for a fixed small time step τ, define a sequence $\{x_k^\tau\}_k$ by $x_{k+1}^\tau \in \arg\min_x F(x) + \frac{|x - x_k^\tau|^2}{2\tau}$

• Importance : • Practical numerical method for approximating the curve.

• Easier generalization to metric space, than $x' = -\nabla F(x)$ itself.

• Properties : • Existence of solution for mild F (e.g. Lipschitz and lower bounded by $C_1 - C_2|x|^2$).

• $\frac{x_{k+1}^\tau - x_k^\tau}{\tau} \in -\partial F(x_{k+1}^\tau)$: implicit Euler scheme (more stable but hard than explicit one : gradient descent).

- Convergence:
  - Define $v_{k+1}^{\tau} \triangleq \dfrac{x_{k+1}^{\tau} - x_k^{\tau}}{\tau}$ and $v^{\tau}(t) = v_{k+1}^{\tau}$, $t \in (k\tau, (k+1)\tau]$.
  - Define two kinds of interpolations:

    1). $x^{\tau}(t) = x_k^{\tau}$, $t \in (k\tau, (k+1)\tau]$.

    2). $\hat{x}^{\tau}(t) = x_k^{\tau} + (t - k\tau) v_{k+1}^{\tau}$, $t \in (k\tau, (k+1)\tau]$.

  - $\hat{x}^{\tau}$ is continuous and $(\hat{x}^{\tau})' = v^{\tau}$
  - x^{τ} is not continuous, but $v^{\tau}(t) \in -\partial F(x^{\tau}(t))$.

- Thm: If $F(x_0) < +\infty$, $\inf F(x) > -\infty$, then up to a subsequence $\tau_j \to 0$, both $\hat{x}^{\tau_j}$ and x^{τ_j} converge uniformly to a same curve $x \in H^1(\mathbb{R}^n)$ and v^{τ_j} weakly convergence in $L^2(\mathbb{R}, \mathbb{R}^n)$ to a vector function v s.t. $x' = v$ and

  1). $v(t) \in \partial F(x(t))$ a.e. if F is λ-convex

  2). $v(t) = -\nabla F(x(t))$, $\forall t$ if F is C^1.

- Details:
  - 1. L^p space. ...
  - 2. Weak convergence in a Hilbert space $\mathcal{H}$.
  - 3. $H^k(\Omega)$ space ($\Omega \subset \mathbb{R}^n$).
  - 4. ....

• Characterizing Properties

• Motivation : • $x' = -\nabla F(x)$ (or $x' \in -\partial F(x)$) is hard to generalize to metric space. There is nothing but distance in metric space, so $\nabla F(x)$ or $\partial F(x)$ cannot be defined! (different from manifold).

• Use two properties of gradient flow that can characterize it and can be generalized to metric space.

• Two charactering properties of gradient flow in $\mathbb{R}^d$:

• Engergy Dissipation Equality (EDE) for $F \in C^1(\Omega)$, $\Omega \subset \mathbb{R}^n$:

$$F(x(s)) - F(x(t)) = \int_s^t \left(\frac{1}{2} |x'(r)|^2 + \frac{1}{2} |\nabla F(x(r))|^2 \right) dr, \ \forall \ 0 \le s < t < 1.$$

is equivalent to $x' = -\nabla F(x)$. Note it is equivalent even for "$\geq$" (i.e. "$\geq$" $\Longleftrightarrow$ "$=$").

• Evolution Variational Inequality (EVI) for λ-convex function F :

$$\frac{d}{dt} \frac{1}{2} |x(t) - y|^2 \le F(y) - F(x(t)) - \frac{\lambda}{2} |x(t) - y|^2, \ \forall y \in X.$$

• Sometimes also denoted as EVI_λ

• It is important for establishing the uniqueness and stability of gradient flow.

Gradient Flow in Metric Space

Generalization of Basic Concepts

- For metric space (X, d),

- Def : Metric derivative of a curve $w: [0,1] \to X$

$$|w'|(t) = \lim_{h \to 0} \frac{d(w(t+h), w(t))}{h} \quad , \text{ if the limit exists.}$$

- If w is Lipschitz, $|w'|(t)$ exists for a.e. $t \in [0,1]$.

- $d(w(t_0), w(t_1)) \leq \int_{t_0}^{t_1} |w'|(s)\, ds.$

- For metric space (X, d), In (X, d), w' cannot be defined, but $|w'|$ can.

- Def : $w: [0,1] \to X$ is absolutely continuous if $\exists g \in L^1([0,1])$ s.t.

$$d(w(t_0), w(t_1)) \leq \int_{t_0}^{t_1} g(s)\, ds, \quad \forall t_0 < t_1.$$

  Let $AC(X)$ be the set of such curves.

- AC $\Rightarrow$ Lipschitz ;
- AC $\Rightarrow$ Metric derivative exists a.e.

- Def : Length of curve $w: [0,1] \to X$:

$$\mathrm{Length}(w) := \sup \left\{ \sum_{k=0}^{n-1} d(w(t_k), w(t_{k+1})) : n \geq 1,\ 0 = t_0 < \dots < t_n = 1 \right\}.$$

- If $w \in AC(X)$, $\mathrm{length}(w) = \int_0^1 |w'|(t)\, dt.$

• Def [Geodesic]: between x_0 and x_1 in X: a curve w s.t. $w(0) = x_0$, $w(1) = x_1$.

and $Length(w) = \min_{\hat{w}} \{ Length(\hat{w}) : \hat{w}(0) = x_0, \hat{w}(1) = x_1 \}$.

• This is the generalization of straight lines in $\mathbb{R}^n$, and is used to extend convexity.

• Def : • Length space: metric space (X, d) s.t.

$$\forall x, y \in X,\ d(x,y) = \inf_{w \in AC(X)} \{ Length(w) : w(0) = x,\ w(1) = y \}.$$

• Geodesic space: length space and geodesic exists for any pair of points.

• Riemann manifolds are geodesic spaces.

• For geodesic space (X, d),

• Def : [Geodesic Convexity]: in a geodesic metric space, a function $F : X \to \mathbb{R}$ that is convex

along geodesics: $F(x(t)) \le (1-t) F(x(0)) + t F(x(1))$.

where $x(t)$ is a geodesic joining $x(0)$ and $x(1)$.

• Def [λ-geodesic convexity]: in a geodesic metric space, a function $F : X \to \mathbb{R}$ that is λ-convex

along geodesics: $F(x(t)) \le (1-t) F(x(0)) + t F(x(1)) - \lambda \frac{t(1-t)}{2} d^2(x(0), x(1))$

- Def : • $g : X \to \mathbb{R}$ is an upper gradient of $F : X \to \mathbb{R}$: for every Lipschitz curve x,

$$|F(x(0)) - F(x(1))| \le \int_0^1 g(x(t))\, |x'|(t)\, dt.$$

  - Local Lipschitz constant of F :

$$|\nabla F|(x) = \limsup_{y \to x} \frac{|F(x) - F(y)|}{d(x,y)}.$$

  - Descending slope (slope) of F :

$$|\nabla^- F|(x) = \limsup_{y \to x} \frac{[F(x) - F(y)]_+}{d(x,y)}.$$

- If F is Lipschitz , $|\nabla F|$ is an upper gradient.

Generalization of Gradient Flow to Metric Space.

- Three ways to generalize gradient flow to metric space : EDE-GF , EVI-GF, MMS-GF.

- Def [EDE-GF]: Let (X, d) be a metric space, $F : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ is an upper gradient of F. EDE-GF is a curve $x : [0,1] \to X$ with metric derivative a.e. s.t.

$$F(x(s)) - F(x(t)) = \int_s^t \left(\frac{1}{2}|x'(r)|^2 + \frac{1}{2} g(x(r))^2 \right) dr, \quad \forall\ 0 \le s < t < 1.$$

- Existence is easy to guarantee

- Not enough to guarantee uniqueness.

• Def [EVI-GF]: Let (X,d) be a geodesic space, $F: X \to \mathbb{R}$ is λ-geodesically convex.

EVI-GF is a curve $x:[0,1] \to X$ s.t.

$$\frac{d}{dt}\frac{1}{2}d(x(t),y)^2 \le F(y) - F(x(t)) - \frac{\lambda}{2}d(x(t),y)^2, \quad \forall y \in X.$$

• A strong condition; existence is hard to guarantee

• Def [Generalized MMS]: Generalization of Minimizing Movement Scheme in a metric space (X,d): for Lipschitz $F: X \to \mathbb{R} \cup \{+\infty\}$, define

$$x_{k+1}^{\tau} \in \arg\min_{x} F(x) + \frac{d(x, x_k^{\tau})^2}{2\tau}.$$

• Define two kinds of interpolations in a similar way:

1). $x^{\tau}(t) = x_k^{\tau}$, $t \in (k\tau, (k+1)\tau]$.

2). $\hat{x}^{\tau}(t) = x_k^{\tau} + (t-k\tau)v_{k+1}^{\tau}$, $t \in (k\tau, (k+1)\tau]$ to be the constant-speed geodesic between x_k^{τ} and x_{k+1}^{τ}.

• Def [Constant-speed geodesic]: in a length space, a curve $w: [t_0, t_1] \to X$ s.t.

$$d(w(t), w(s)) = \frac{|t-s|}{t_1 - t_0} \cdot d(w(t_0), w(t_1)), \quad \forall t, s \in [t_0, t_1].$$

- Constant-speed geodesics are geodesics: $\text{Length}(w) = \int_{t_0}^{t_1} \frac{d(w(t_0), w(t_1))}{t_1 - t_0} dt = d(w(t_0), w(t_1))$.

- The following are equivalent: ① $w : [t_0, t_1] \to X$ is a constant-speed geodesic joining x_0 and x_1.

② $w \in AC(X)$ and $|w'|(t) = \frac{d(w(t_0), w(t_1))}{t_1 - t_0}$ a.e.

③ $w \in \arg\min \left\{ \int_{t_0}^{t_1} |w'|(t)^p \, dt : w(t_0) = x_0,\ w(t_1) = x_1 \right\}$, $\forall p > 1$.

- Define v^τ. On metric (length) spaces, only its the norm can be defined: Set $|v^\tau|$ as the piecewise constant speed of $\hat{x}^\tau$,

$$|v^\tau|(t) = d(x^\tau_{k+1}, x^\tau_k)/\tau, \quad t \in (k\tau, (k+1)\tau].$$

- Def [MMS-GF]: Let (X, d) be a metric space (not necessarily length space).

A curve $x : [0, T] \to X$ is called MMS-GF if there exists a sequence $\tau_j \to 0$ s.t. x^{τ_j} uniformly converges to x in (X, d).

Table: Conclusion of extentions of gradient flow to metric space

Extension	Requirement	Existence	Uniqueness and Contractivity
EVI-GF	X geodesic space, F λ-geod. convex	Hard. C^2G^2 is a sufficient condition	Guaranted
EDE-GF	X metric space	Easy	Not guaranteed
MMS-GF	X metric space	Relatively easy. "$\{x : F(x) \le c\}$ compact and F Lipschitz" or "F λ-geod. convex" suffices	Not guaranteed

- EVI-GF $\subset$ EDE-GF
- MMS-GF $\subset$ EDE-GF if "$\{x : F(x) \le c\}$ compact, F Lipschitz, F and $|\nabla^- F|$ lower-semicont., $|\nabla^- F|$ is an upper grad. of F" or "F λ-geod. convex"

Gradient Flows on Wasserstein Spaces

Recap. of Optimal Transport Pbms.

• Settings: Let X, Y be two measurable spaces, $\mu \in P(X)$ and $\nu \in P(Y)$ are fixed measures.

Let $c: X \times Y \to \mathbb{R}$ be a cost function.

• Def [push-forward of a measure] For a measurable function $T: X \to Y$ and a measure $\mu \in P(X)$, define the push-forward of μ under T, $T\#\mu$, to be a measure on Y s.t.

$$T\#\mu(A) = \mu(T^{-1}(A)), \quad \forall A \in \sigma\text{-algebra of } Y.$$

• e.g. For $X = Y = \mathbb{R}^n$ and T invertible, then in terms of p.d.f., $T\#\mu = (\mu \circ T^{-1}) |\det(\nabla T^{-1})|$.

• Monge's pbm: $\quad (MP) \quad \inf\limits_{T\#\mu = \nu} \int_X c(x, T(x))\, d\mu(x).$

- Optimal T is called a (optimal) transport map.
- The problem may not be feasible.

• Kantorovich's pbm: $\quad (KP) \quad \inf\limits_{\gamma \in \Pi(\mu,\nu)} \int_{X \times Y} c(x,y)\, d\gamma(x,y).$

where $\Pi(x,y) := \{\gamma \mid (\pi_X)_\# \gamma = \mu, \ (\pi_Y)_\# \gamma = \nu\}.$

- Optimal γ is called a optimal transport plan.
- The problem is always feasible.

• MP is a special case of KP, where γ is restricted to the form $\gamma = (id \times T)_\#\mu$. If T^* exists, $\gamma^* = (id \times T^*)_\#\mu$ is also optimal.

- Dual Kantorovich Pbm:

    - Direct form: (DKP) $\sup_{\substack{\phi \in L^1(X),\ \psi \in L^1(Y) \\ \phi(x)+\psi(y) \leq c(x,y)}} \int_X \phi \, d\mu + \int_Y \psi \, d\nu.$

- Reformulation:

- Def: • c-transform (c-conjugate) of $\chi : X \to \overline{\mathbb{R}}$, $\chi^c : Y \to \overline{\mathbb{R}}$ is defined as

$$\chi^c(y) := \inf_{x \in X} c(x,y) - \chi(x).$$

    - $\Psi_c(X) := \{ \chi^c \mid \chi : X \to \overline{\mathbb{R}} \}$. $\psi : Y \to \overline{\mathbb{R}}$ is c-concave if $\psi \in \Psi_c(X)$.

$$(DKP'): \sup_{\phi \in \Psi_c(X)} \int_X \phi \, d\mu + \int_Y \phi^c \, d\nu.$$

- Def [Kantorovich potential]: The optimal ϕ of (DKP') is called Kantorovich potential denoted by φ.

- When c is uniformly continuous (e.g. when c is continuous and X is compact), then the existence of Kantorovich potential φ can be proven (by AA thm).

- Remark: Strong duality holds: $KP(\mu,\nu) = DKP(\mu,\nu)$.

• Dual Kantorovich Pbm :

• Special case 1. : $X = Y$, $c(x,y) = d(x,y)$ is a distance.

$$(DKP1) \quad \sup_{\phi \in Lip_1} \int_X \phi \, d\mu - \int_X \phi \, d\nu .$$

• Special case 2 : $X = Y = \Omega \subset \mathbb{R}^n$ and $c(x,y) = \frac{1}{2}|x-y|^2$:

• Thm : • for quadratic cost and $\Omega \subset \mathbb{R}^n$ close, bounded and connected, exists a unique optimal transport plan γ^* for (KP).

• Additionally, if μ is absolutely continuous, optimal transport map T^* exists and $\gamma^* = (id, T^*)\#\mu$. Moreover, there exists a Kantorovich potential φ s.t. $\nabla\varphi$ is unique μ-a.e. and $T = \nabla u$ a.e. where $u(x) := \frac{x^2}{2} - \phi(x)$ is convex.

• Cor : • Under the same condition, any gradient of a convex function is an optimal map between μ and its image measure.

• Optimal transport map uniquely exists for $c(x,y) = h(x-y)$ with h strictly convex.

(e.g. $|x-y|^p$, $p > 1$).

The Wasserstein Space :

• Def : On metric space (X, d), for $p \geq 1$ and a fixed point $x_0 \in X$, define

$m_p(\mu) := \int_X d(x, x_0)^p d\mu(x)$, and $\mathcal{P}_p(X) := \{\mu \in \mathcal{P}(X) : m_p(\mu) < +\infty\}$, which is independent

of the choice of x_0.

• Thm : $W_p(\mu, \nu) := \left[\inf_{\gamma \in \Pi(\mu,\nu)} \int_X d(x,y)^p d\gamma(x,y) \right]^{1/p}$ is a distance on $\mathcal{P}_p(X)$.

• Def [Wasserstein space] : $\mathbb{W}_p(X) := (\mathcal{P}_p(X), W_p)$.

• Thm : In $\mathbb{W}_p(X)$ with $p \geq 1$, given $\mu, \mu_n \in \mathcal{P}_p(X)$, $n \in \mathbb{N}$, the following are equivalent :

- $\mu_n \to \mu$ w.r.t. W_p ;
- $\mu_n \to \mu$ and $m_p(\mu_n) \to m_p(\mu)$;
- $\int_X \phi \, d\mu_n \to \int_X \phi \, d\mu$, $\forall \phi \in \{\phi \in C^0(X) : \exists A, B \in \mathbb{R}$ s.t. $|\phi(x)| \leq A + B\, d(x, x_0)^p, \forall x, x_0 \in X\}$.

• Thm [McCann's displacement interpolation].

- If $\Omega \in \mathbb{R}^d$ is convex, then $\mathbb{W}_p(\Omega)$ is a length space, and for $\mu, \nu \in \mathbb{W}_p(\Omega)$ and γ as optimal transport plan from μ to ν, then $\mu^\gamma(t) := (\pi_t)_\# \gamma$, where $\pi_t(x,y) := (1-t)x + ty$, is a constant-speed geodesic.

- If $p > 1$, then all the constant-speed geodesics are of this form.

- If additionally μ is absolutely continuous, then there is only one geodesic, whose form is

  $\mu_t = (T_t)_\# \mu$, where $T_t := (1-t)\,id + tT$.

  where T is the optimal transport map from μ to ν.

- Geodesic convexity in $\mathbb{W}_2(\Omega)$.

- Def [Important functionals on $\mathbb{W}_2(\Omega)$]

  For $f: \mathbb{R} \to \mathbb{R}$ convex, $V: \Omega \to \mathbb{R}$, $W: \mathbb{R}^d \to \mathbb{R}$ symmetric $(W(-x) = W(x))$.

  define $\mathcal{F}(\rho) = \int f(\rho(x))\,dx$, $\mathcal{V}(\rho) = \int V(x)\,d\rho$, $\mathcal{W} = \frac{1}{2}\iint W(x-y)\,d\rho(x)\,d\rho(y)$.

- Thm:
  - λ-convexity on Ω of V (or W) $\Rightarrow$ λ-geodesic convexity on $\mathbb{W}_2(\Omega)$ of $\mathcal{V}$ (or $\mathcal{W}$).
  - $f(0) = 0$ and $s^d f(s^{-d})$ is convex and decreasing, Ω is convex, $1 < p < \infty$

    $\Rightarrow$ $\mathcal{F}$ is geodesic convex in $\mathbb{W}_2(\Omega)$.

Gradient Flows on $\mathbb{W}_2(\Omega)$, $\Omega \subset \mathbb{R}^n$

- Continuity equation: $\mathbb{W}_p(\Omega)$: it is of probability distributions. The curve / flow / dynamics in $\mathbb{W}_p(\Omega)$, μ_t, represents the evolution of distributions. This evolution can be associated with (viewed as a result of) an evolution / dynamics in $\mathbb{R}^n$, represented by vector field v_t.

The typical relation between them is the continuity equation:

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0.$$

- Thm: Let $p>1$. $\Omega \subset \mathbb{R}^d$ open, bounded and connected.
  - Let $\{\mu_t\}_{t\in[0,1]}$ be an AC curve in $\mathbb{W}_p(\Omega)$. Then for a.e. $t \in [0,1]$ there exists a vector field $v_t \in L^p(\mu_t; \mathbb{R}^d)$ s.t.

    1). $\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0$ is satisfied in the sense of distributions.

    2). for a.e. $t \in [0,1]$, $\|v_t\|_{L^p(\mu_t)} \leq |\mu'|(t)$.

  - Conversely, if $\{\mu_t\}_{t\in[0,1]} \subset \mathcal{P}_p(\Omega)$ and $\forall t$ we have a vector field $v_t \in L^p(\mu_t; \mathbb{R}^d)$ with $\int_0^1 \|v_t\|_{L^p(\mu_t)} dt < +\infty$ solving $\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0$, then $\{\mu_t\}_{t\in[0,1]}$ is AC in $\mathbb{W}(\Omega)$ for a.e. $t \in [0,1]$, $|\mu'|(t) \leq \|v_t\|_{L^p(\mu_t)}$.

  - $\Rightarrow \|v_t\|_{L^p(\mu_t)} = |\mu'|(t)$.

- Let $F: \mathbb{W}_2(\Omega) \to \overline{\mathbb{R}}$ be a functional on $\mathbb{W}_w(\Omega)$. define the gradient flow w.r.t F:

$$\rho_{k+1}^{\tau} \in \arg\min_{\rho} F(\rho) + \frac{W_2^2(\rho, \rho_k^{\tau})}{2\tau} \quad \cdots\cdots \text{ MMS-GF.}$$

- General existence conditions apply e.g. $\{\rho : F(\rho) \leq c\}$ compact and F Lipschitz, or F λ-geodesically convex.

- Thm: Let $F: \mathbb{W}_2(\Omega) \to \overline{\mathbb{R}}$ be λ-geodesically convex, the MMS-GF w.r.t. F exists. Let ρ_t^0, ρ_t^1 be two solutions, and define $E(t) := \frac{1}{2} W_2^2(\rho_t^0, \rho_t^1)$. Then $E(t) \leq e^{-\lambda t} E(0)$, which implies uniqueness for a given initial condition, and stability and exponential convergence for $\lambda > 0$.

- To relate F and the vector field v_t, we need the notion of first variation.

- Def [First Variation]: First Variation of a functional $G: P(\Omega) \to \mathbb{R}$ is defined as $\frac{\delta G}{\delta \rho}(\rho) : \Omega \to \mathbb{R}$ s.t. $\frac{d}{d\varepsilon} G(\rho + \varepsilon \chi)\big|_{\varepsilon=0} = \int \frac{\delta G}{\delta \rho}(\rho)(x)\, d\chi(x)$, $\forall \chi \in \{\chi : \exists \varepsilon_0 \text{ s.t } \forall \varepsilon \in [0, \varepsilon_0], \rho + \varepsilon \chi \in P(\Omega)\}$.

- Thm: $\mathcal{F}(\rho) = \int f(\rho(x))\, dx$, $\frac{\delta \mathcal{F}}{\delta \rho} = f'(\rho)$.

$$\mathcal{V}(\rho) = \int V(x)\, d\rho, \quad \frac{\delta \mathcal{V}}{\delta \rho} = V$$

$$\mathcal{W} = \iint W(x-y)\, d\rho(x)\, d\rho(y), \quad \frac{\delta \mathcal{W}}{\delta \rho} = W * \rho \quad \text{(convolution).}$$

(Thm)

- The first variation of Wasserstein distance with cost function c: $\frac{\delta W_c(\rho, \nu)}{\delta \rho} = \varphi$,

if ρ, ν are defined on $\Omega \subset \mathbb{R}^d$, $c: \Omega \times \Omega \to \mathbb{R}$ continuous, and Kantorovich potential

φ is unique and c-concave.

- Thm: The Minimizing Movement Scheme $\rho_{k+1}^{\tau} \in \arg\min_{\rho} F(\rho) + \frac{W_2^2(\rho, \rho_k^{\tau})}{2\tau}$, the optimality condition

is $\frac{\delta F}{\delta \rho}(\rho_{k+1}^{\tau}) + \frac{\varphi}{\tau} = \text{Const}.$

where φ is the Kantorovich potential from ρ_{k+1}^{τ} to ρ_k^{τ}.

- Relation between T^* and φ: $T^*(x) = x - \nabla\varphi$

Relation between v_t and T: $v_t(x) = (x - T(x))/\tau$.

So in the limit $\tau \to 0$, the gradient flow w.r.t. F induces a flow in $\mathbb{R}^n$:

$$v_t(x) = -\nabla\left(\frac{\delta F}{\delta \rho}(\rho_t)\right)(x),$$

and the flow ρ_t in $\mathbb{W}_2(\Omega)$ is:

$$\partial_t \rho_t - \nabla \cdot \left(\rho_t \nabla\left(\frac{\delta F}{\delta \rho}(\rho_t)\right)\right) = 0.$$