2-Wasserstein Distance W , $W = W(F_A, F_B) = \left(\int_0^1 |F_A^{-1}(u) - F_B^{-1}(u)|^2 du \right)^{1/2}$.

$$d = W^2 = d(F_A, F_B) = \int_0^1 |F_A^{-1}(u) - F_B^{-1}(u)|^2 du = \underbrace{(\mu_A - \mu_B)^2}_{\text{location}} + \underbrace{(\sigma_A - \sigma_B)^2}_{\text{size}} + \underbrace{2\sigma_A\sigma_B(1 - \rho^{A,B})}_{\text{shape}}$$

$\rho^{A,B}$: Pearson correlation.

- JKO flows. let $F: \mathbb{R}^n \to \mathbb{R}$ be convex. want to minimize $F(x)$;

Gradient flow: $\begin{cases} x'(t) = -\nabla F(x(t)) \\ x(0) = x_0 \end{cases}$

Discretize in Time via Backward Euler.

$$\frac{x^{n+1} - x^n}{\tau} = -\nabla F(x^{n+1}).$$

$$\Rightarrow \quad \frac{x^{n+1} - x^n}{\tau} + \nabla F(x^{n+1}) = 0.$$

$$\Rightarrow \quad \nabla\left(\underbrace{\frac{|x - x^n|^2}{2\tau}}_{\text{convex}} + \underbrace{F(x)}_{\text{convex}} \right)\Bigg|_{x = x_{n+1}} = 0.$$

$$\Rightarrow \quad x^{n+1} \in \arg\min \left\{ \frac{|x - x^n|^2}{2\tau} + F(x) \right\}.$$

can define a scheme like this on a metric space. (X, d).

let $F: X \to \mathbb{R}$. be l.s.c and bdd below.

- Define $x_\tau^{n+1} \in \arg\min \left\{ F(x) + \frac{d(x, x^n)^2}{2\tau} \right\}$.

- interpolate to all t.

e.g. $x_\tau(t) = x_\tau^n$ if $t \in ((n-1)\tau, n\tau]$

study limit as $\tau \to 0$.

X: probability measures. and d : Wasserstein 2 measures.

Consider : $F : P(\Omega) \to \mathbb{R}$ and $d = W_2$.

- Ω is compact
- F is l.s.c and bdd below.

we previously use the continuity equation. $\rho_t + \nabla \cdot (\rho v) = 0$ to "flow" densities.

Goal : find velocity field v s.t. this flow agrees with $\lim_{T \to 0} X_T(t)$.

- Investigate optimality condition in the JKO scheme.

we need to compute the 1st variation

we want to perturb $\rho \in P(\Omega)$. to $\rho + \varepsilon \chi$

Need $\rho + \varepsilon \chi \in P(\Omega)$ so that $F(\rho + \varepsilon \chi)$ is well define.

restrict to χ s.t. $\sigma = \rho + \chi \in P(\Omega)$ $\forall \varepsilon > 0$. small ε.

$\Rightarrow \rho + \varepsilon \chi = \rho + \varepsilon(\sigma - \rho) = \rho(1 - \varepsilon) + \varepsilon \sigma \in P(\Omega)$ as long as $\rho, \sigma \in P(\Omega)$.

$\forall \sigma \in P(\Omega) \cap L^\infty_2(\Omega)$.

The first variation of F, $\frac{\partial F}{\partial \rho}(\rho)$ is such that.

$$\frac{d}{d\varepsilon} F(\rho + \varepsilon \chi)\Big|_{\varepsilon = 0} = \int \frac{\delta F}{\partial \rho}(\rho) \chi(x) \, dx. \quad \forall \chi = \sigma - \rho.$$

$\sigma \in P(\Omega) . \cap L^\infty_c(\Omega)$

$$\int \chi(x)\,dx = \int \sigma - \rho\,dx = 1-1=0.$$

$$\int \left(\frac{\partial F}{\partial \rho} + c\right)\chi(x)\,dx = \int \frac{\partial F}{\partial \rho}\chi(x)\,dx + c\int \chi(x)\,dx$$

The 1^{st} variation is defined uniquely only up to addivitive constants.

Come back to $G(\rho) = F(\rho) + \dfrac{W_2^2(\rho, \rho_\tau^n)}{2\tau}$.

we need to compute $\dfrac{\delta G}{\delta \rho}(\rho) = \dfrac{\delta F}{\delta \rho}(\rho) + \dfrac{1}{2\tau}\cdot\dfrac{\delta W_2^2}{\delta \rho}(\rho, \rho_\tau^n)$

We have the dual formulation: $W_2^2(f,g) = 2 \inf\limits_{\pi\in\Pi(f,g)} \int \dfrac{|x-y|^2}{2}\,d\pi(x,y).$

$$= 2\cdot \max_{u,v}\left\{ \int u f\,dx + \int v g\,dy \,\middle|\, u(x)+v(y) \le \tfrac{1}{2}|x-y|^2 \right\}.$$

$$= 2\cdot \max_{u}\left\{ \int u f\,dx + \int u^c g\,dy \right\}.$$

$$\frac{d}{d\varepsilon} W_2^2(f+\varepsilon\chi, g)\Big|_{\varepsilon=0}.$$

$$= 2\cdot\frac{d}{d\varepsilon}\max_u\left\{\int u(f+\varepsilon\chi)\,dx + \int u^c g\,dy\right\}\Big|_{\varepsilon=0}.$$

$$= 2\cdot\int u^*\chi\,dx.$$

where u^* achieves the max in the previous line. i.e. u^* is the potential association with the cost $\frac{1}{2}|x-y|^2$.

when using the OT, the optimal map is

$$T(x) = x - \nabla u^*(x). = x - (\nabla h)^{-1}(\nabla u^*(x)) \quad \text{where } h(z) = \tfrac{1}{2}|z|^2.$$

$$\Rightarrow \quad \frac{\delta W_2^2}{\partial \rho}(\rho, \rho_\tau^n) = \partial u^*.$$

$T(x) = x - \nabla u^*(x)$ is the optimal map from ρ to ρ_τ^n.

The JKO scheme is $\quad \rho_\tau^{n+1} = \arg\min \left\{ F(\rho) + \frac{W_2^2(\rho, \rho_\tau^n)}{\tau} \right\}.$

$$= \arg\min \ G(\rho)$$

$$\Rightarrow \quad \frac{\delta G}{\delta \rho}(\rho_\tau^{n+1}) + C = 0.$$

$$\Rightarrow \quad \frac{\delta F}{\delta \rho}(\rho_\tau^{n+1}) + \frac{u^*}{\tau} = \text{Constant}.$$

$$\Rightarrow \quad 0 = \nabla\left(\frac{\partial F}{\partial \rho}\right) + \frac{\nabla u^*}{\tau}$$

$$= \nabla\left(\frac{\partial F}{\partial \rho}\right) + \frac{x - T(x)}{\tau}.$$

$$\Rightarrow \quad \underbrace{\frac{T(x) - x}{\tau}}_{} = \nabla\left(\frac{\partial F}{\partial \rho}\right)$$

velocity! of a flow from ρ_τ^{n+1} to ρ_τ^n.

the flow we want should have velocity

$$V(x) = -\frac{T(x) - x}{\tau} = -\nabla\left(\frac{\partial F}{\partial \rho}\right).$$

This is the velocity associated with our time-discrete scheme.

If everything works out as $\tau \to 0$, we expect our JKO scheme to limit to this flow.

$$\rho_t + \nabla \cdot (\rho V) = 0.$$

or $\rho_t - \nabla\cdot\left(\rho \frac{\delta F}{\delta \rho}\right) = 0.$ This is the PDE associate with gradient of in the W_2 metric.

e.g. $F(\rho) = \int \rho \cdot \log \rho \, dx$ (negative entropy) we want a flow that maximizes entropy

$$\frac{d}{d\varepsilon} F(\rho + \varepsilon \chi)\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon} \int (\rho + \varepsilon \chi) \log(\rho + \varepsilon \chi)\, dx \Big|_{\varepsilon=0}$$

$$= \int (\chi \log \rho + \chi)\, dx$$

$$= \frac{\partial}{\partial \varepsilon} \int \rho \log(\rho + \varepsilon \chi) + \varepsilon \chi \log(\rho + \varepsilon \chi)\, dx$$

$$= \int \frac{\rho \cdot \chi}{\rho + \varepsilon \chi} + \chi \log(\rho + \varepsilon \chi) + \varepsilon \chi \frac{\chi}{\rho + \varepsilon \chi} \Big|_{\varepsilon = 0} dx$$

$$= \int \chi + \chi \log(\rho)\, dx.$$

$$\Rightarrow \frac{\partial F}{\partial \rho} = \log \rho + 1.$$

$$\nabla\left(\frac{\partial F}{\partial \rho}\right) = \nabla(\log \rho + 1) = \frac{1}{\rho} \cdot \nabla \rho.$$

$\Rightarrow$ the Gradient flow is

$$0 = \rho_t - \nabla \cdot \left(\rho \cdot \frac{1}{\rho} \cdot \nabla \rho\right)$$

$$= \rho_t - \nabla \cdot (\nabla \rho) = \rho_t - \Delta \rho.$$

$\Rightarrow \rho_t = \Delta \rho$ (heat equation).

• e.g. $F(\rho) = \int \rho \cdot \log \rho \, dx + \int V(x) \cdot \rho \, dx$

$\Rightarrow \rho_t - \Delta \rho - \nabla \cdot (\rho \nabla V) = 0.$ (Fokker-Planck).

• e.g. $F(\rho) = \frac{1}{m-1} \int \rho^m \, dx.$

$\Rightarrow \rho_t - \Delta(\rho^m) = 0.$ (porous medium).

• e.g. $F(\rho) = \int \rho \log \rho - \frac{1}{2} \int |\nabla u_\rho|^2.$ where $-\Delta u_\rho = \rho.$

$$\Rightarrow \begin{cases} \rho_t + \nabla \cdot (\rho \cdot \nabla u) - \Delta \rho = 0 \\ -\Delta u = \rho. \end{cases}$$ (Keller-Segel, chemotaxis)

• e.g. $F(\rho) = \frac{1}{2} \iint W(x-y) \, d\rho(x) \, d\rho(y).$ W: potential.

$\Rightarrow \rho_t - \nabla \cdot (\rho ((\nabla W) * \rho)) = 0$ (aggregation model.).