Dynamic Optimal Transport

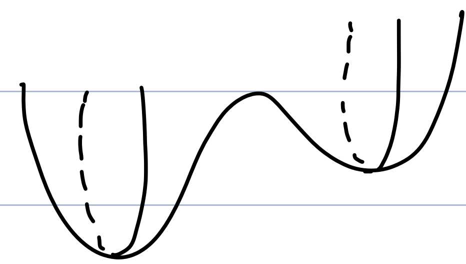
© Katy Craig, 2025

Plan

Dynamic 2-Wasserstein Metric

⇝ geometry

⇝ gradient flows



Generalizations of W_2

1. Graph Wasserstein Metric
2. Hellinger Kantorovich Metric
3. Vector-Valued Optimal Transport

References

[AGS] Ambrosio, Gigli, Savaré, Gradient Flows, 2005

[M] Maas, "Gradient flows of the entropy", 2011

[EM] Erbar, Maas, "Ricci Curvature...", 2012

[GLM] Gangbo, Li, Mou, "Geodesic of Minimal...", 2017

Dynamic 2-Wasserstein Metric

Given $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$

Kantorovich:

$$W_2^2(\mu,\nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \int |x_1 - x_2|^2 \, d\gamma$$

Dynamic (Benamou-Brenier):

$$W_2^2(\mu,\nu) = \inf_{(\rho_t(x),\, v_t(x))} \int_0^1 \overbrace{\int_{\mathbb{R}^d} |v_t(x)|^2 \, d\rho_t(x)}^{\text{Kinetic energy}} \, dt$$

continuity eqn (weak sense)

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$$

$$\rho_0 = \mu, \quad \rho_1 = \nu$$

$v_t(x)$ tells which direction $\rho_t(x)$ goes at time t, location x

Ex: $\rho_0 = \frac{1}{2} \mathbb{1}_{[-1,1]}(x)\,dx$, $v_t(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$

$$\int_0^1 \int_{\mathbb{R}} |v_t(x)|^2 \, d\rho_t(x)\, dt = 1 = W_2^2(\rho_0, \rho_1)$$

<u>Key properties of dynamic formulation:</u>

- the minimum is achieved by geodesics
- if there exists an OT map T_μ^ν, the the unique minimum is

$$\rho_t = (\underbrace{(1-t)x + t\,T_\mu^\nu(x)}_{=:T_t(x)})\# \mu$$

$$v_t(x) = T_\mu^\nu(T_t^{-1}(x)) - T_t^{-1}(x)$$

Geodesics on Metric Spaces

Given a metric space $(\mathcal{X}, d)$, a curve $x : [0,1] \to \mathcal{X}$ is a (constant speed) geodesic if $d(x_t, x_s) = |t-s| d(x_0, x_1)$, $\forall t, s \in [0,1]$.

Geometry:

$\boxed{\text{Thm (AGS)}}$ Given $\rho : [0,1] \to \mathcal{P}_2(\mathbb{R}^d)$ smooth in time, $\exists!\, v$ s.t. (ρ, v) satisfy cty eqn and $v_t \in \overline{\{\nabla \varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}$ for a.e. t.

$$\overline{\{\nabla \varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}} = \mathrm{Tan}_{\rho_t} \mathcal{P}_2(\mathbb{R}^d)$$

This leads to a heuristic Riemannian structure...

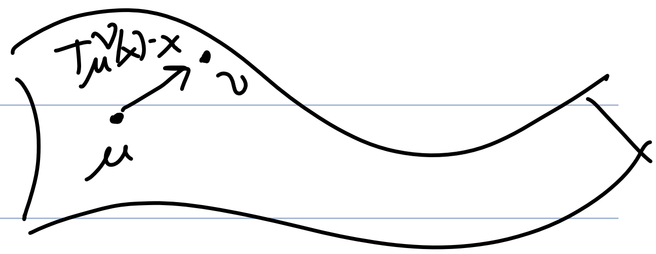
the unique v_t guaranteed by AGS

$$W_2^2(\mu,\nu) = \inf_{\substack{\rho:[0,1]\to \mathcal{P}_2(\mathbb{R}^d) \text{ smooth in time} \\ \rho_0=\mu,\ \rho_1=\nu}} \int_0^1 \langle v_t, v_t\rangle_{\rho_t}\, dt$$

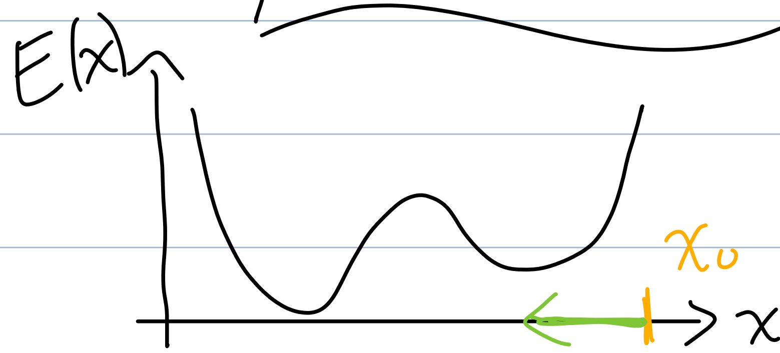
"Riemannian metric" $\langle v,u\rangle_\rho = \int v(x)\cdot u(x)\, d\rho$

Utility of perspective, e.g. logarithmic map:

$$\mathrm{Log}_\mu(\nu) = T_\mu^\nu(x) - x$$



Gradient Flows:



Given $E: \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ smooth, $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$, what smooth curve $\rho: [0,1] \to \mathcal{P}_2(\mathbb{R}^d)$ makes E decrease as quickly as possible?

Finite Dimensional Warmup:

Consider $E: \mathbb{R}^d \to \mathbb{R}$ smooth

Gradient descent: $\begin{cases} \dot{x}_t = -\nabla E(x_t) \\ x_t\big|_{t=0} = x_0 \end{cases}$

Why is this the best path?

$$\frac{d}{dt}E(x_t) = \nabla E(x_t)\cdot \dot{x}_t = -|\nabla E(x_t)|^2$$

Cauchy Schwarz guarantees this is the ~~best~~ best choice of $\dot{x}_t$ for making $\frac{d}{dt}E(x_t)$ as small as possible (up to normalization of magnitude.

Functional derivative

Given $E: L^2(\mathbb{R}^d) \to \mathbb{R}$, $f \in L^2(\mathbb{R}^d)$, if $\exists\ \varphi \in L^2(\mathbb{R}^d)$ s.t.

$$\lim_{h \to 0} \frac{E(f+hg) - E(f)}{h} = \int_{\mathbb{R}^d} \varphi g \, dx, \quad \forall g \in L^2(\mathbb{R}^d),$$

then we define $\frac{\delta E}{\delta f} := \varphi$.

Note: If $d\rho_t = \rho_t(x)dx$ for $\rho_t(x) \in L^1 \cap L^\infty$, then $\rho_t \in L^2(\mathbb{R}^d)$.

What is the best curve to decrease E as fast as possible?

$$\frac{d}{dt} E(\rho_t) = \lim_{h \to 0} \frac{E(\rho_{t+h}) - E(\rho_t)}{h}$$

$$= \lim_{h \to 0} \frac{E(\rho_t + h\dot{\rho}_t + O(h^2)) - E(\rho_t)}{h}$$

$$= \int_{\mathbb{R}^d} \frac{\delta E}{\delta \rho_t} \dot{\rho}_t \, dx$$

$$\partial_t \rho_t = -\nabla \cdot (v_t \rho_t)$$

$$= -\int_{\mathbb{R}^d} \frac{\delta E}{\delta \rho_t} \nabla \cdot (v_t \rho_t)$$

$$= \int_{\mathbb{R}^d} \nabla \frac{\delta E}{\delta \rho_t} \cdot v_t \rho_t$$

$$= \left\langle \nabla \frac{\delta E}{\delta \rho_t}, v_t \right\rangle_{\rho_t}$$

Best choice of $v_t = -\nabla \frac{\delta E}{\delta \rho_t}$

General heuristic PDE for 2-Wasserstein GFs:

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

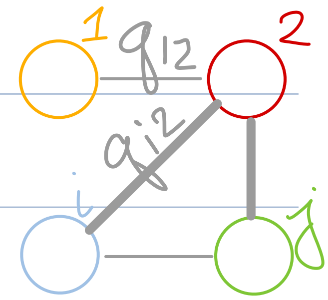
Ex: $E(\rho) = \int \rho(x) \log(\rho(x)) dx$

$$\partial_t \rho_t = \Delta \rho_t$$

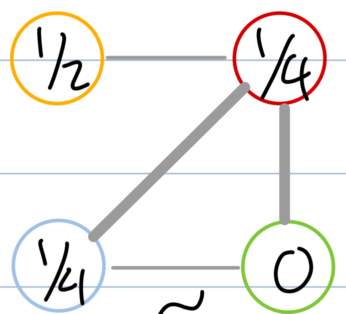
Graph Wasserstein Metric

Let G denote an n node, connected, symmetric, weighted graph.

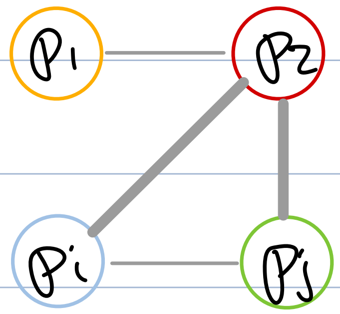
$q_{ij} = q_{ji}$



Let $p, \tilde{p} \in \mathcal{P}(G) = \{ p \in \mathbb{R}^n_+ : \sum_{i=1}^{n} p_i = 1 \}$



$\tilde{p}$



p

Graph operators

Given $\phi \in \mathbb{R}^n$, $\nabla_G \phi = [\phi_i - \phi_j]_{i,j=1}^{n} \in M_{n \times n}$

Given $v \in M_{n\times n}$, $\mathrm{div}_G(v) = \left[-\frac{1}{2} \sum_{j=1}^{n} (v_{ij} - v_{ji}) q_{ij} \right]_{i=1}$

Key facts:

- $\langle \nabla_G \phi, v \rangle_G = - \langle \phi, \mathrm{div}_G v \rangle_{\mathbb{R}^n}$
- $\mathrm{div}_G \nabla_G \phi = \Delta_G \phi$

Dynamic Wassenstein metric: $\boxed{M}$

$$W_G^2(\rho,\tilde{\rho}) = \inf_{(\rho_t,v_t)} \frac{1}{2}\sum_{i,j=0}^{n}\int_0^1 |v_{ij,t}|^2 \check{\rho}_{ij,t}\, q_{ij}\, dt$$

$$\partial_t \rho_t + \mathrm{div}_G(v_t \check{\rho}_t) = 0$$

$$\rho_0 = \rho,\ \rho_1 = \tilde{\rho}$$

where $\check{\rho} = [\check{\rho}_{ij}]_{i,j=1}^{n} = [\theta(\rho_i,\rho_j)]_{i,j=1}^{n}$

Examples of θ:

① $\theta(\rho_i,\rho_j) = \frac{\rho_i+\rho_j}{2}$, ② $\theta(\rho_i,\rho_j) = \sqrt{\rho_i\rho_j}, \dots$

Role of edge weights:

After a change of variables...

$$W_G^2(\rho,\tilde{\rho}) = \inf \frac{1}{2}\sum_{i,j}\int_0^1 |v_{ij,t}|^2 \check{\rho}_{ij,t}\, q_{ij}^{-1}\, dt$$

$$\partial_t \rho_t + \mathrm{div}_{[n]}(v_t \check{\rho}_t) = 0$$

$$\rho_0 = \rho,\ \rho_1 = \tilde{\rho}$$

Geometry:

there is a heuristic Riemannian structure

Thm (M) $(\mathcal{P}_{>0}(G), W_G)$ is a smooth Riemannian manifold.

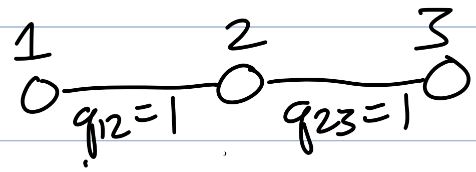
Bad News:

- $(\mathcal{P}_{>0}(G), W_G)$ is not complete; geodesics between points in interior can touch bdy. $\boxed{\text{GLM}}$
- the minimizer of dynamic problem does not necessarily exist but if it exists and is strictly positive, it is a geodesic. (as a curve in the space of probability measures)

For those who want a deep dive into this last statement, here is more detail...

① If one further restricts the constraint set of the dynamic formulation of W_G to only consider solutions of the graph continuity equation satisfying $\rho_t \in \mathcal{P}(G)$ $\forall\, t \in (0,1)$, then minimizers exist. $\boxed{\text{GLM, Thm 4.5}}$

② I believe that the following example should show that the minimizing curve is not always a curve in $\mathcal{P}(G)$. Moreover, understanding the structure of W_G geodesics is important to understanding when the space may be branching, which is relevant in applications of vector-valued OT.

Graph  , $\rho = \begin{bmatrix}\rho_1\\ \rho_2\\ \rho_3\end{bmatrix}$, $\tilde{\rho} = \begin{bmatrix}\tilde{\rho}_1\\ \tilde{\rho}_2\\ \rho_3\end{bmatrix}$

$$\theta(\rho_i,\rho_j) = \frac{\rho_i + \rho_j}{2}$$

<u>Continuity eqn constraint</u>

$$\partial_t \rho_t + \nabla_G \cdot (\check{\rho} v) = 0$$

$$\Updownarrow$$

$$\partial_t \rho_{1,t} - v_{12}\left(\frac{\rho_{1,t}+\rho_{2,t}}{2}\right) = 0$$
$$\partial_t \rho_{2,t} + v_{12}\left(\frac{\rho_{1,t}+\rho_{2,t}}{2}\right) + v_{23}\left(\frac{\rho_{2,t}+\rho_{3,t}}{2}\right) = 0$$
$$\partial_t \rho_{3,t} - v_{23}\left(\frac{\rho_{2,t}+\rho_{3,t}}{2}\right) = 0$$

$$\Updownarrow$$

$q_t = \rho_{1,t} + \rho_{2,t}$, $r_t = \rho_{2,t} + \rho_{3,t}$
$= 1 - \rho_{1,t}$

$$\dot{q}_t + \frac{1}{2} v_{23} r_t = 0$$
$$\dot{r}_t + \frac{1}{2} v_{12} q_t = 0$$

Objective function

$$\int_0^1 \sum_{i,j} v_{ij,t}^2 \check{\rho}_{ij,t} q_{ij} dt = \int_0^1 v_{12,t}^2 q_t + v_{23,t}^2 r_t \, dt$$

applying cty eqn constraint

$$= 4\int_0^1 \frac{(\mathring{r}_t)^2}{q_t} + \frac{(\mathring{q}_t)^2}{r_t} dt$$

$r_t = 1-\rho_{1,t}$, $\mathring{r}_t = -\mathring{\rho}_{1,t}$

$q_t = 1-\rho_{3,t}$, $\mathring{q}_t = -\mathring{\rho}_{3,t}$

$$= 4\int_0^1 \frac{(\mathring{\rho}_{1,t})^2}{1-\rho_{3,t}} + \frac{(\mathring{\rho}_{3,t})^2}{1-\rho_{1,t}} dt$$

By symmetry,

$\rho_{3,t} = \rho_{1,1-t}$

$$= 8\int_0^1 \frac{(\mathring{\rho}_{1,t})}{1-\rho_{1,1-t}} dt$$

This leads to the problem

$$\min_{\rho_1:[0,1]\to\mathbb{R}} 8\int_0^1 \frac{(\mathring{\rho}_{1,t})^2}{1-\rho_{1,1-t}} dt$$

s.t. $\rho_1(0) = \rho_1$

$\rho_1(1) = \tilde{\rho}_1$

This should be solvable (at least numerically??).

Once we solve this, we obtain

$\rho_{3,t} = \rho_{1,1-t}$

$\rho_{2,t} = 1 - \rho_{1,t} - \rho_{3,t}$.

I expect that $\rho_{2,t}$ will be zero for some values of t, which will prove existence of a minimizer that leaves $\mathcal{P}(G)$.