Lecture #15: Optimal Transport

- We now aim to understand the geometry of Wasserstein space, by analyzing smoothness and geodesics.
- We have established that $\mathcal{P}_p(\mathbb{R}^d)$ endowed with the Wasserstein-p metric is indeed a metric space, for $p \geq 0$ (albeit we will often focus on $p \geq 1$ and especially $p=2$.)
- Let us recall curves and notions of "speed" in general metric spaces.

__Defn__: Let (X,d) be a metric space.

- We say a __curve__ is a continuous function $w:[0,1] \to (X,d)$
- The metric derivative at time $t \in (0,1)$ is

$$|w'|(t) := \lim_{h \to 0} \frac{d(w(t+h), w(t))}{|h|}$$

__Theorem__: Let $w:[0,1] \to X$ be Lipschitz (i.e. $\exists L > 0$ s.t. $\forall t_1, t_2 \in [0,1]$, $d(w(t_1), w(t_2)) \leq L \cdot |t_1 - t_2|$). Then $|w'|: [0,1] \to \mathbb{R}_{\geq 0}$ exists a.e. and for $t<s$,

$$d(w(t), w(s)) \leq \int_t^s |w'|(\tau)\,d\tau$$

__Defn__: We say a curve $w:[0,1] \to X$ is __absolutely continuous__ if there exists $g \in L^1([0,1])$ such that

$$d(w(t_0), w(t_1)) \leq \int_{t_0}^{t_1} g(s)\,ds \qquad \text{for all } t_0 < t_1.$$

Let __$AC(X)$__ be the set of absolutely continuous curves on X.

- We can define a notion of length of curve as follows:

__Defn__: Let $w: [0,1] \to X$ be a curve. Define the the __length__ of w as

$$\text{Length}(w) := \sup_{\substack{0<t_1<\dots<t_n \\ n\geq 1}} \sum_{k=0}^{n-1} d(w(t_k), w(t_{k+1}))$$

Note that AC curves have their lengths in terms of the metric derivative:

$$\text{Length}(w) = \int_0^1 |w'|(t)\,dt$$

- With these notions in mind, we can define notions of geodesic for (X,d):

Defn: • A curve $w : [0,1] \to X$ is said to be a geodesic between $x_0, x_1 \in X$ if

$$\text{Length}(w) = \min_{\substack{\tilde{w}:[0,1]\to X \\ w(0)=x_0 \\ w(1)=x_1}} \text{Length}(\tilde{w}).$$

- A space (X,d) is a length space if $d(x,y) = \inf_{\substack{w\in AC(X) \\ w(0)=x \\ w(1)=y}} \text{Length}(w)$

- A space (X,d) is said to be a geodesic space if $d(x,y) = \min_{\substack{w\in AC(X) \\ w(0)=x \\ w(1)=y}} \text{Length}(w)$.

- We say a curve $w : [0,1] \to X$ in a length space is a constant-speed geodesic between $w(0)$ and $w(1) \in X$ if it satisfies

$$d(w(t), w(s)) = |t-s| \cdot d(w(0), w(1)) \quad \text{for all } t,s \in [0,1].$$

Proposition (Equivalences to Constant Speed): Let $p > 1$ and let $w : [0,1] \to X$ be a curve with $w(0) = x_0$, $w(1) = x_1$. TFAE:

(i) w is a constant speed geodesic

(ii) $w \in AC(X)$ and $|w'|(t) = d(w(0), w(1))$ a.e.

(iii) w solves $\min\left\{\int_0^1 |\tilde{w}'|(t)^p\,dt\right\}$, where the min is over curves with $\tilde{w}(0) = x_0$, $\tilde{w}(1) = x_1$

- We are interested in understanding curves and geodesics in $\mathcal{W}_p = (\mathcal{P}_p(\mathbb{R}^d), W_p)$, the <u>Wasserstein-$p$ space.</u> To do so, we want to make a connection with the so-called <u>continuity equation</u>

- Imagine we have particles in $\mathbb{R}^d$ that evolve in time (variables x and t, respectively). For each (x,t) pair, we can ask "what is the density of particles in location x at time t"? This quantity will be measured by the density $\rho(t,x)$. Consider a velocity field $V = \{V_t(x)\}$, which act on a particle x via the ODE

$$\begin{cases} y_x'(t) = V_t(y_x(t)), \\ y_x(0) = x, \end{cases}$$

interpreting $y_x(t)$ as "the position of x at time $t \geq 0$"

- We can characterize the dynamical behavior of $\rho(t,x)$ as follows. Let $\rho_0(x) := \rho(0,x)$ be an initial particle distribution. Let $Y_t(x) = y_x(t)$ and let $\rho_t := (Y_t)_\# \rho_0$. We can think of ρ_t as the distribution of particles at time t, which are evolved from initial distribution ρ_0 according to the vector field $V = \{V_t\}$.

- Crucially, $V = \{V_t\}$ and ρ_t are related as

$$\partial_t \rho_t + \nabla \cdot (\rho_t V_t) = 0 \qquad \text{(Continuity Equation)}$$

- We will connect this equation to curves in Wasserstein space, but first we will investigate it a little in its own right.

<u>Definition</u>: • We say pairs of measures and vector fields $\{(\rho_t, v_t)\}_{t\in[0,T]}$ for $\Omega \subseteq \mathbb{R}^d$ (bounded or all of $\mathbb{R}^d$)

$v_t \in L^1(\rho_t;\mathbb{R}^d)$ with $\int_0^T \|v_t\|_{L^1(\rho_t)}\, dt = \int_0^T \int_\Omega \underbrace{|v_t|}_{\text{usual norm on } \mathbb{R}^d} d\rho_t\, dt < \infty$

<u>Satisfy the continuity equation on $(0,T)$ ~~in~~ in the distributional sense</u>, if for any bounded and Lipschitz $\phi \in C^1_c((0,T)\times\bar{\Omega})$, we have

$$\int_0^T\int_\Omega [\partial_t \phi]\, d\rho_t\, dt + \int_0^T\int_\Omega \nabla\phi\cdot v_t\, d\rho_t\, dt = 0$$

• Note, ~~...~~ This implicitly imposes, where Ω is bounded, the no-flux condition $\rho_t v_t\cdot n = 0$ for normal vectors n on $\partial\Omega$. We can also impose conditions on initial and final distributions ρ_0, ρ_T via replacing the RHS above with

$$\int_\Omega \phi(T,x)\, d\rho_T(x) - \int_\Omega \phi(0,x)\, d\rho_0(x).$$

• We say that $\{(\rho_t,v_t)\}_{t\in[0,T]}$ as above <u>Solves the continuity equation in the weak sense</u> if for any $\psi \in C^1_c(\bar{\Omega})$, the function

$$t \mapsto \int \psi\, d\rho_t$$

is absolutely continuous and for a.e. t,

$$\frac{d}{dt}\int_\Omega \psi\, d\rho_t = \int_\Omega \nabla\psi\cdot v_t\, d\rho_t.$$

<u>Proposition (distributional and weak solutions coincide for CE)</u>: Every weak solution to CE is a distributional solution and every distributional solution admits an a.e. unique

weak solution.

Proof: We recall that separable functions of the form $\phi(t,x) = a(t)\psi(x)$ are such that their linear combinations (finite) are dense in $C^1([0,T]\times K)$ for all K compact.

Now, let $\{(\rho_t, v_t)\}$ be a distributional solution. Let $\phi(t,x) = a(t)\psi(x)$ be a separable solution for $\psi \in C^1_c(\bar{\Omega})$, $a \in C^1_c((0,1))$. Then by assumption of distributional solution,

$$\int_0^T a'(t)\int_\Omega \psi(x)\,d\rho_t\,dt + \int_0^T a(t)\int_\Omega \nabla\psi\cdot v_t\,d\rho_t\,dt = 0$$

Since this holds for all a, this implies the distributional time derivative of

$$t \mapsto \int_\Omega \psi(x)\,d\rho_t$$

is equal to $\displaystyle\int_\Omega \nabla\psi\cdot v_t\,d\rho_t$, i.e. $\displaystyle\frac{d}{dt}\int_\Omega \psi\,d\rho_t = \int_\Omega \nabla\psi\cdot v_t\,d\rho_t$,

as desired.

On the other hand, we see that a weak separable solution is a distributional solution. By density, we have that all weak solutions are distributional solutions. ∎

//

We want to connect the continuity equation to the "flow" of the vector field $\{v_t\}_{t\in[0,T]}$. Let us recall how this works.

Proposition: Consider the ODE
$$\begin{cases} y_x'(t) = v_t(y_x(t)), & t\in[0,T] \\ y_x(0) = x \end{cases}$$

i) If v_t is continuous for every t as a function of x ($v_t : \mathbb{R}^d$), then the above ODE admits a solution on a neighborhood of $t=0$. If moreover

$$|V_t(x)| \leq C_0 + C_1|x| \quad \text{for all } t, \text{ some constants } C_1, C_2,$$

then the solution exists globally.

(ii) If V_t is Lipschitz in X uniformly in t, then letting $L = \sup Lip(V_t)$, we have that the solution is unique and defines a flow $Y_t(x) := y_x(t)$ That is Lipschitz in X, invertible, and has bi-Lipschitz constant ~~exp~~ $\exp(tL)$.

<u>Next time</u>: Properties of CE and connection to Wasserstein spaces.