"Dynamical" view of the OT distance W_2

- Domain (X, d) Polish metric space
  ↳ will later assume it geodesic, or $= \mathbb{R}^d$, or Riemannian manifold
- W_2 distance : $W_2^2(\mu, \nu) = \min_{\pi \in \Gamma(\mu,\nu)} \mathbb{E}_{(X,Y)\sim\pi}\left[d^2(X, Y)\right]$

PLAN :

$\boxed{A}$ Metric structure

$\boxed{B}$ Geodesic structure

$\boxed{C}$ Differential structure — Benamou-Brenier formula

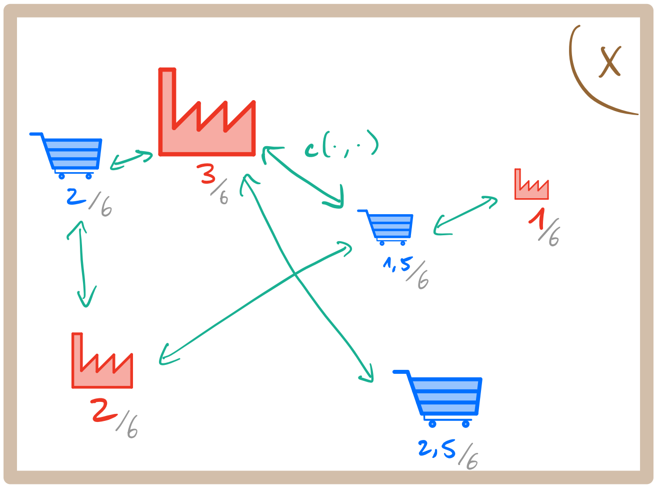
$\boxed{D}$ Riemannian formalism — Otto calculus

NOT covered :
- $c = d^p$, $p \neq 2$
- regularity of stuff (besides absolute continuity (AC))
- topological stuff (completeness, compactness, relat° betw cvgce notions...)
- proofs, rigor ...
- convexity of the Kantorovich potentials

Refs : all of Figalli's 2021 book (nice & short)

all of Ambrosio's 2021 book (good tradeoff completeness / niceness)

$\boxed{0}$ The OT problem with general cost function



 = factory, production

$\rightsquigarrow \ \mu \in \mathcal{P}(X)$

 = shops, consumption

$\rightsquigarrow \ \nu \in \mathcal{P}(X)$

$c(\underset{\in X}{x}, \underset{\in X}{y})$ = cost of transporting one unit of commodity from x to y

Want: transport plan $\big(\pi(\cdot|x)\big)_{x \in \mathrm{support}(\mu)}$ s.t $\int \pi(\cdot|x)\, d\mu(x) = \nu$

and $\int c(x,y)\ \pi(y|x) d\mu(x)$ low.

Optimal transport cost :

$$W_c(\mu,\nu) = \min_{\pi \in \mathcal{P}(X \times X)} \int_{X \times X} c(x,y)\, d\pi(x,y)$$

$$\text{s.t} \begin{cases} \int_X d\pi(\cdot,y) = \mu \\ \int_X d\pi(x,\cdot) = \nu \end{cases} \quad \text{i.e.} \quad \begin{matrix} (x)_\# \pi = \mu \\ (y)_\# \pi = \nu \end{matrix}$$

Def.

$\rightsquigarrow$ Linear programming,

Kantorovich duality,

Kantorovitch potentials ...

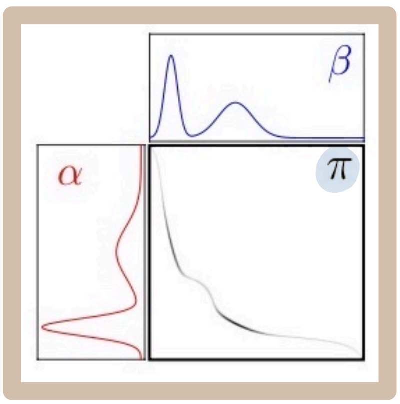
NOT covered

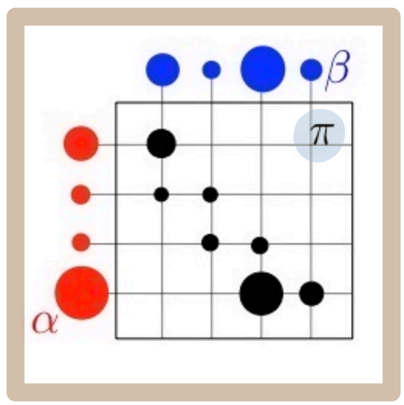
A Metric structure

Prop : (X, d) Polish metric space $\Longrightarrow (\mathcal{P}_2(X), W_2)$ Polish metric space

Proof : uninteresting

$$W_2^2(\mu,\nu) = \min_{\pi \in \mathcal{P}(X\times X)} \int_{X\times X} d^2(x,y)\, d\pi(x,y) \quad \text{s.t.} \quad \begin{cases} (x)_\# \pi = \mu \\ (y)_\# \pi = \nu \end{cases}$$





Examples :

- $X = \mathbb{R}^d$, d = Euclidean dist
- $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$, d = Euclidian dist

Refreshers ?

$\xrightarrow{\text{will need}}$ Monge formulation of W_2 (as $\inf_T$)

$\xrightarrow{\text{will need}}$ $\mu_0 \ll$ Leb $\Rightarrow T^*$ exists (Brenier thm)

$\xrightarrow{\text{might}}$ Kantorovich duality, Kantorovitch potentials

$\xrightarrow{\text{won't}}$ topology of narrow cvgce (a.k.a in duality with $\mathcal{C}_b(X)$)

[A+21, lec. 9 & 10]
[F+21, § 3.1]

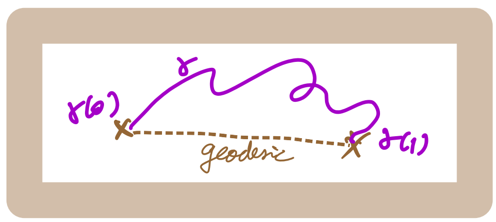
$\boxed{B}$ Geodesic structure

Prop: (X,d) geodesic space $\Rightarrow$ $(\mathcal{P}_2(X), W_2)$ geodesic space

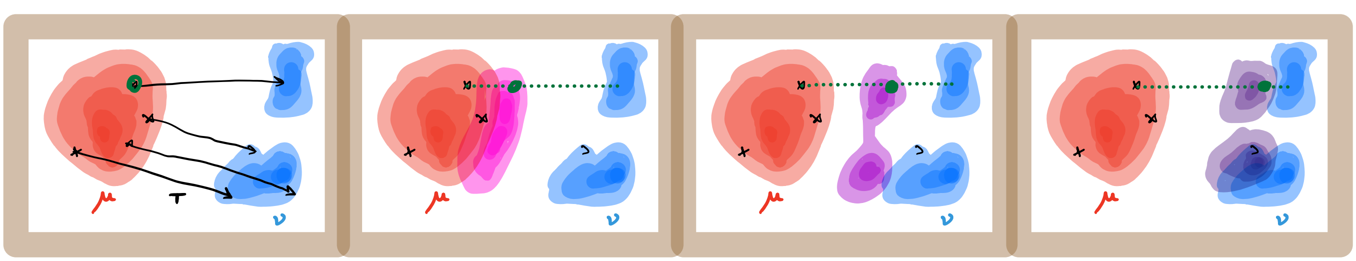
Reminders (?) and defs about geodesic spaces:

- $\gamma : [0,1] \to X$ is called AC if $\exists g \in L^1([0,1])$; $d(\gamma(s),\gamma(t)) \leq \int_s^t g$,
  $AC([0,1], X)$ a.k.a. "a-e differentiable & satisfies F.Th.Calc"
- can define metric derivative $|\gamma'(t)| = \lim_{h\to 0} \frac{d(\gamma(t+h),\gamma(t))}{h}$ for a.e t
- Length: $\ell(\gamma) = \int_0^1 |\gamma'(t)|\,dt \geq d(\gamma(0),\gamma(1))$
- γ is called a geodesic if $\ell(\gamma) = d(\gamma(0),\gamma(1))$

and a constant-speed geodesic if $|\gamma'(t)| = cst.$



Prop: Let π^* optimal coupling of μ, ν.
Then $\mu_t = (\Sigma_t)_\# \pi^*$ where $\Sigma_t(x,y) = (1-t)x + ty$ — can be generalized to geo. spaces
defines a constant-speed geodesic in $(\mathcal{P}_2(X), W_2)$. In particular that space is geodesic.



μ_0 $\mu_{0,25}$ $\mu_{0,5}$ $\mu_{0,75}$

Rk: If there exists an optimal transport map T^* from μ to ν, then we can choose $\pi^* := (id \times T)_\# \mu$.

(actually one can show π^* is anyway unique)

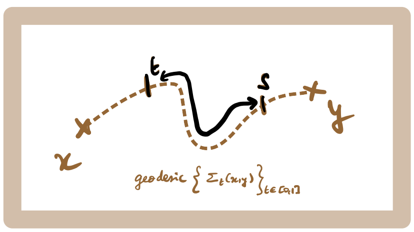
Then, $\mu_t = (\Sigma_t)_\# \pi^* = \big((1-t)x + t\,T(x)\big)_\# \mu$.

Proof: To bound $W_2^2(\mu_s, \mu_t) = \min_{\bullet \in \Gamma(\mu_s,\mu_t)} \int \|x-y\|^2 \, d\bullet(x,y)$,

argument generalizes to geo. spaces

consider the coupling: $(\Sigma_s, \Sigma_t)_\# \pi^*$.

$$W_2^2(\mu_s,\mu_t) \le \int \|x-y\|^2 \, d(\Sigma_s,\Sigma_t)_\#\pi^*(x,y)$$



$$= \int \|\Sigma_s(x,y) - \Sigma_t(x,y)\|^2 \, d\pi^*(x,y)$$

$$= (t-s)^2 \int \|x-y\|^2 d\pi^*(x,y)$$

$$= (t-s)^2 \, W_2^2(\underset{\mu}{\underset{\parallel}{\mu_0}}, \underset{\nu}{\underset{\parallel}{\mu_1}}).$$

One can show that

$$\Big[\forall s,t \in [0,1],\ W_2(\mu_s,\mu_t) \le |t-s|\, W_2(\mu_0,\mu_1)\Big] \Rightarrow (\mu_t)_t \text{ constant-speed geodesic.} \ \blacksquare$$

$=$, by triangle inequality!

Prop: Under some add'tl assumptions on (X,d), e.g. for $X = \mathbb{R}^d$, all constant-speed geodesic curves are as described above.

That is, for any constant-speed geodesic $(\mu_t)_{t\in[0,1]}$, we have

$$\mu_t = (\Sigma_t)_\# \pi^* \quad \text{where } \pi^* \in \Gamma_{\text{optimal}}(\mu_0,\mu_1).$$

Alternative equivalent presentation "à la Ambrosio" :

- Action : $\mathcal{A}_2(\gamma) = \int_0^1 |\gamma'(t)|^2 dt$

Prop: $\mathcal{A}_2(\gamma) \geqslant \ell(\gamma)^2 \geqslant d(\gamma(0), \gamma(1))^2$
and γ cst-speed geodesic $\Longleftrightarrow$ $\mathcal{A}_2(\gamma) = d(\gamma(0), \gamma(1))^2$

Thm: $$W_2^2(\mu, \nu) = \boxed{\min_{\eta \in \mathcal{P}(\mathcal{C}([0,1],X))} \int_{\mathcal{C}([0,1],X)} \mathcal{A}_2(\gamma)\, d\eta(\gamma) \quad \text{s.t.} \quad (e_0)_\# \eta = \mu, \ (e_1)_\# \eta = \nu}$$

Any optimal η induces a constant-speed geodesic in $(\mathcal{P}_2(X), W_2)$ via $\mu_t = (e_t)_\# \eta$.

$\boxed{C}$ Differential structure

Prop: (X, d) Riemannian manifold $\Longrightarrow$ $(\mathcal{P}_2(X), W_2)$ has a Riemannian-like structure

Throughout this section, $X = \mathbb{R}^d$ with Euclidian dist.

C.I) The continuity equation (CE)

Consider a (time-depdt) vector field $v : [0, T) \times \mathbb{R}^d \to \mathbb{R}^d$, and

$$\boxed{\frac{dx(t)}{dt} = v_t(x(t)) \quad , \quad x(0) = x_0} \qquad \text{(ODE)}$$

Prop: Let a r.v. $X_0 \sim \mu_0$ and X_t the position after following (ODE) for t seconds. Then $\mathrm{Law}(X_t)$ satisfies (CE) below:

$\boxed{\text{i.e.}}$

Let $X(t, x)$ the flow associated to (ODE).

For any μ_0, $\mu_t := (X(t, \cdot))_{\#} \mu_0$ satisfies (CE):

$$\boxed{\partial_t \mu_t = -\operatorname{div}(v_t \mu_t) \quad , \quad \mu_0 \text{ specified}} \qquad \text{(CE)}$$

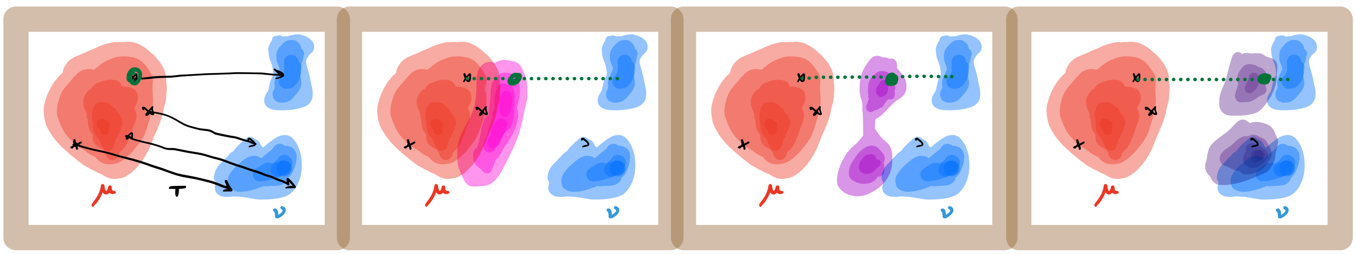
distributionally, i.e.

$$\forall \varphi \in \mathcal{C}^\infty(\mathbb{R}^d), \quad \frac{d}{dt}\left(\int \varphi \, d\mu_t\right) = \int (\nabla \varphi)^T v_t \, d\mu_t .$$

Proof: Let $\varphi \in \mathcal{C}^\infty(\mathbb{R}^d)$. By definition $\mu_t := \mathrm{Law}(X_t)$ is given by

$$
\begin{aligned}
\frac{d}{dt}\int \varphi \, d\mu_t &= \frac{d}{dt}\mathbb{E}\,\varphi(X_t) \\
&= \mathbb{E}\left[\nabla\varphi(X_t)^T \frac{dX_t}{dt}\right] \\
&= \mathbb{E}\left[\nabla\varphi(X_t)^T v_t(X_t)\right] = \int [\nabla\varphi]^T v_t \, d\mu_t . \qquad \square
\end{aligned}
$$

C.II) The Benamou-Brenier formula



Reinterpret this drawing: consider the flow followed by each "particle".

I.e.,
- Fix μ and ν, supp. $\exists$ optimal map T
- We know $\mu_t := \big[\underbrace{(1-t)x + t\,T(x)}_{=:\,T_t}\big]_\# \mu$ is a geodesic $\mu \to \nu$
- μ_t can also be intepreted as:
  - draw particle $X_0 \sim \mu_0 = \mu$
  - make it evolve as $\dfrac{dX_t}{dt} = T(X_0) - X_0$ $\longrightarrow$ clearly $X_t = T_t(X_0)$

    $= (T - id)(T_t^{-1}(X_t))$.

$\Rightarrow$ there exists a vector field $v_t^* := (T - id) \circ T_t^{-1}$ that induces μ_t as the solution to (CE).

(∃ by convexity of Kantorovich potentials → [F+21, p.24])

This v_t^* must be "optimal" among all such v_t's.

in a sense not straightforward to figure out

Thm: $$W_2^2(\bar\mu_0,\bar\mu_1) = \min_{(\mu_t, v_t)_t} \int_0^1 \left[\int_{\mathbb{R}^d} \|v_t\|^2 \, d\mu_t\right] dt$$

s.t. $\mu_0=\bar\mu_0$, $\mu_1=\bar\mu_1$, $v_t \in L^2(\mu_t)$ $\forall t$,

and $\partial_t \mu_t = -\operatorname{div}(v_t \mu_t)$. ← (CE)

Proof in the case where $\exists T$ optimal transport map from $\bar\mu_0$ to $\bar\mu_1$:

($\geq$) Consider, as in the discussion above,

$$\mu_t^* = \big(\underbrace{(1-t)x + tT(x)}_{=: T_t}\big)_\# \bar\mu_0, \qquad v_t^* = \nabla T_t \circ T_t^{-1}.$$

It will be easy to check from the calculations below that
$\int_0^1 \int_{\mathbb{R}^d} \|v_t^*\|^2 \, d\mu_t^* \, dt = W_2^2(\bar\mu_0,\bar\mu_1)$.

($\leq$) Let any admissible $(\mu_t, v_t)_t$.

Let $X(t,x)$ the flow associated to (ODE) with velocity field v_t. By uniqueness of the solution for (CE), we have $\mu_t = (X(t,\cdot))_\# \bar\mu_0$.

$\leadsto$ $X(1,\cdot)$ is a valid transport map from $\bar\mu_0$ to $\bar\mu_1$. So

$$\begin{aligned}
W_2^2(\bar\mu_0,\bar\mu_1) &\leq \int_{\mathbb{R}^d} \|X(1,x) - x\|^2 \, d\bar\mu_0(x) \\
&= \int_{\mathbb{R}^d} \left\| \int_0^1 \frac{d}{dt} X(t,x) \, dt \right\|^2 d\bar\mu_0(x) \\
&\leq \int_{\mathbb{R}^d} \int_0^1 \Big\| \underbrace{\frac{d}{dt} X(t,x)}_{= v_t(X(t,x)) \text{ by (ODE)}} \Big\|^2 dt \, d\bar\mu_0(x) \\
&= \int_{\mathbb{R}^d} \int_0^1 \|v_t(x')\|^2 \, dt \, \underbrace{d(X(t,\cdot))_\# \bar\mu_0(x')}_{= d\mu_t(x')}. \quad \blacksquare
\end{aligned}$$

In the case where there is no optimal transport map: proof more confusing to write, but same idea. [A+21, §17.1]

Recap: for $(X,d) = (\mathbb{R}^d, \|\cdot\|)$, we know that $(\mathcal{P}_2(\mathbb{R}^d), W_2)$

A $\rightsquigarrow$ is a metric space

B $\rightsquigarrow$ is a geodesic space, and the unique constant-speed geodesic from μ to ν is $\mu_t = (\Sigma_t)_\# \pi^*$ where $\Sigma_t(x,y) = (1-t)x + ty$, $\{\pi^*\} = \Gamma_{optimal}(\mu,\nu)$

$= [(1-t)x + tT(x)]_\# \mu_0$ if $\pi^* = (id \times T)\mu$

$\rightsquigarrow$ by general property of geod. spaces,

$$W_2^2(\mu,\nu) = \min_{\gamma \in AC([0,1], \mathcal{P}_2(\mathbb{R}^d))} \int_0^1 \underbrace{|\gamma'(t)|^2}_{\text{(abstract) metric derivative}} dt$$ ("action" of curve γ)

C $\rightsquigarrow$ $$W_2^2(\mu,\nu) = \min_{(\mu_t,v_t)} \int_0^1 \underbrace{\left[\int_{\mathbb{R}^d} \|v_t\|^2 d\mu_t\right]}_{=: \|\mu_t'\|^2} dt \quad \text{s.t. (CE)}$$

just a notation; not even standard

$\Rightarrow$ Natural q.: Does this abstract metric derivative arise from some more concrete structure?

E.g., can $\|\mu_t'\|^2$ be intepreted as $\left\| \frac{d}{dt}\mu_t \right\|^2_{\mu_t}$ in the sense of Riemannian geometry? « $(\mu_t, \frac{d}{dt}\mu_t) \in T\mathcal{M}$ »

[F+21 §4.2]

$\boxed{D}$ Riemannian formalism

Ansatz : $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ behaves "like" a Riemannian manifold $(\mathcal{M}, g)$.

For an AC curve $(\mu_t)_t \subset \mathcal{P}_2(\mathbb{R}^d)$, $\left(\mu_t, \frac{d\mu_t}{dt}\right) \in T\mathcal{M}$ and

$$\mathcal{A}_2((\mu_t)_t) = \inf_{(v_t)_t} \int_0^1 \int_{\mathbb{R}^d} \|v_t\|^2 d\mu_t \, dt \quad \text{s.t} \quad (CE) .$$

It turns out that this suffices to formally derive g :

① $$\mathcal{A}_2((\mu_t)_t) = \inf_{(v_t)_t} \int_0^1 \int_{\mathbb{R}^d} \|v_t\|^2 d\mu_t \, dt \quad \text{s.t} \quad \partial_t \mu_t = -\operatorname{div}(v_t \mu_t)$$

$(\mu_t)_t$ fixed
so constraints on $(v_t)_t$
are separable

$$= \int_0^1 \underbrace{\left\{ \inf_{v \in L^2(\mu_t)} \int_{\mathbb{R}^d} \|v\|^2 d\mu_t \quad \text{s.t} \quad \partial_t \mu_t = -\operatorname{div}(v \mu_t) \right\}}_{=: \|\partial_t \mu_t\|^2_{\mu_t}} dt$$

$$\Rightarrow \forall s \in T_\mu \mathcal{M}, \quad \|s\|_\mu^2 = \inf_{v \in L^2} \int_{\mathbb{R}^d} \|v\|^2 d\mu \quad \text{s.t} \quad s = -\operatorname{div}(v\mu)$$

② Can simplify this : one can show that [F+21, p. 84]

- any minimizer v must be of the form $v = \nabla \phi$ for some $\phi : \mathbb{R}^d \to \mathbb{R}$,
- and the functional eq° in ϕ $\quad -\operatorname{div}(\nabla\phi \cdot \mu) = s \quad$ has a unique solution for any μ, s (smooth enough).

$\Rightarrow \forall s \in T_\mu \mathcal{M}, \; \|s\|_\mu^2 = \int_{\mathbb{R}^d} \|\nabla \phi\|^2 d\mu$ where ϕ solves $-\mathrm{div}(\nabla\phi \cdot \mu) = s$

$$=: \int_{\mathbb{R}^d} \|\nabla(Ls)\|^2 d\mu .$$

Def $\boxed{-\mathrm{div}(\nabla(Ls)\mu) = s}$

③ By parallelogram rule : $\langle a,b\rangle = \dfrac{\|a+b\|^2 - \|a\|^2 - \|b\|^2}{2}$, we get

$$\forall s, s' \in T_\mu \mathcal{M}, \quad \langle s, s'\rangle_\mu = \int_{\mathbb{R}^d} \nabla(Ls)^T \nabla(Ls')\, d\mu .$$

④ Can simplify this :

$$\langle s, s'\rangle_\mu = \int \nabla(Ls)^T \nabla(Ls')\, d\mu$$

$$= - \int Ls \; \underbrace{\mathrm{div}(\nabla(Ls')d\mu)}_{s'}$$

$$= \int (Ls) \times s' = \int (Ls') \times s . \Rightarrow \text{"}L \text{ is symm. for } L^2 \text{ i.p."}$$

Def / Notation (Otto calculus) :

For any $s, s' : \mathbb{R}^d \to \mathbb{R}$ s.t $\underbrace{\int_{\mathbb{R}^d} s = \int_{\mathbb{R}^d} s' = 0}^{\text{since } \mu_t \in \mathcal{P}(\mathbb{R}^d) \Rightarrow \int_{\mathbb{R}^d} (\mu_{t+dt} - \mu_t) = 0}$,

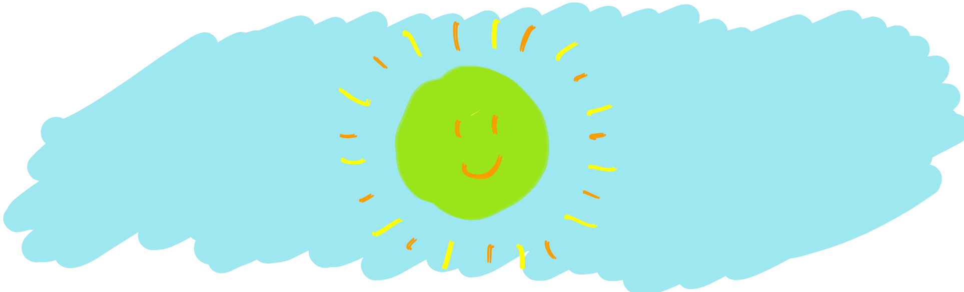
$$\langle s, s'\rangle_\mu := \int_{\mathbb{R}^d} \nabla(Ls)^T \nabla(Ls')\, d\mu$$

where $Ls = \phi$ is the unique solution of $-\mathrm{div}(\nabla\phi \cdot \mu) = s$

$$= \langle \underset{\mathcal{M}(\mathbb{R}^d)}{s}, \underset{\mathcal{E}(\mathbb{R}^d)}{gs'} \rangle_{\mathcal{E}(\mathbb{R}^d)} \quad \text{where} \quad g = -L^T \mathrm{div}(\nabla(L\cdot)\mu) = L .$$

Rk: L depends on μ .

At this point one could check that the metric / geodesic considerations developed above fit in perfectly and that life is beautiful and nature is amazing.



D.I) W_2 Gradient Flow

Def: Let $F : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. We say $F'_\mu \in \mathcal{E}(\mathbb{R}^d)$ is the *first variation* of F at μ if
$\forall \nu \in \underbrace{T_\mu \mathcal{M}}_{\text{i.e. } \nu \in \mathcal{M}(\mathbb{R}^d),\ \int_{\mathbb{R}^d} d\nu = 0}$, $\forall \delta \in \mathbb{R}$,

$$F(\mu + \delta\nu) - F(\mu) = \delta \langle \underset{\mathcal{E}(\mathbb{R}^d)}{\underset{\cap}{F'_\mu}}, \underset{\mathcal{M}(\mathbb{R}^d)}{\underset{\cap}{\nu}} \rangle_{\mathcal{E}(\mathbb{R}^d)} + o(\delta).$$

Prop: Let $F : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ s.t. F'_μ exists.
Then,

$$\mathrm{grad}_W F(\mu) = g^{-1} F'_\mu = L^{-1} F'_\mu = -\mathrm{div}(\nabla(F'_\mu) \cdot \mu).$$

Proof: take an AC curve $(\mu_t)_{t \in [-1,1]}$ in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and compute $\frac{d}{dt}\Big|_{t=0} F(\mu_t)$ in two ways: with the def of F'_μ, and with def of $\mathrm{grad}_W F$. $\square$

Def: W_2-GF is $\boxed{\partial_t \mu_t = -\mathrm{grad}_W F(\mu_t) = +\mathrm{div}(\nabla(F'_{\mu_t}) \cdot \mu_t)}$

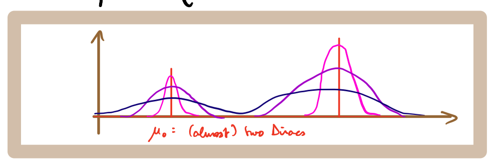
Example : W_2-GF of entropy is heat flow

Let $H[\mu] = \begin{cases} \int_{\mathbb{R}^d} \rho \log \rho \, dx & \text{if } \mu \ll Leb \text{ and } d\mu = \rho\, dx \\ +\infty & \text{otw} . \end{cases}$

Then
$$H'_\mu(x) = \log \rho(x)$$
$$\nabla H'_\mu(x) = \frac{\nabla \rho(x)}{\rho(x)}$$
$$\mathrm{grad}_W H[\mu] = -\mathrm{div}\left(\nabla H'_\mu * \mu\right) = -\mathrm{div}\left(\frac{\nabla \rho}{\rho} * \rho\, dx\right)$$
$$= -\mathrm{div}\left(\nabla \rho\right)$$
$$= -\Delta \rho .$$

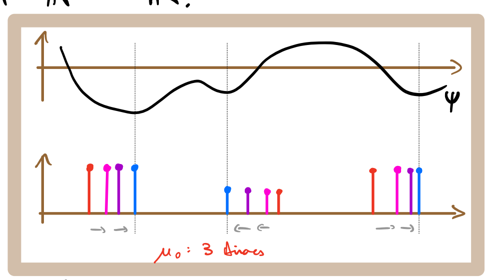
So W_2-GF of H is $\boxed{\partial_t \rho_t = \Delta \rho_t}$ $\quad (d\mu_t = \rho_t \, dx)$

heat equation



Example : potential energy functional

Let $V[\mu] = \int_{\mathbb{R}^d} \Psi \, d\mu$ for some $\Psi : \mathbb{R}^d \to \mathbb{R}$.



Then
$$V'_\mu(x) = \Psi(x)$$
$$\nabla V'_\mu(x) = \nabla \Psi(x)$$
$$\mathrm{grad}_W V[\mu] = -\mathrm{div}\left(\nabla \Psi * \mu\right)$$

So W_2-GF of V is $\quad \partial_t \mu_t = +\mathrm{div}\left(\nabla \Psi * \mu_t\right)$

Equivalently : $\quad X_0^{(l)} \sim \mu_0$

i.e. GF for Ψ $\quad \left(\frac{d X_t^{(l)}}{dt} = -\nabla \Psi(X_t^{(l)}) \right.$

$\forall l, \quad \mu_t = Law(X_t^{(l)})$ $\quad$ by the equivalence (ODE) $\longleftrightarrow$ (CE)

$\leadsto$ simultaneous GF, "no interaction"

Example: relative entropy to fixed target measure

Let $H(\mu | \mu^*) = \begin{cases} \int_{\mathbb{R}^d} \log \frac{d\mu}{d\mu^*} \, d\mu & \text{if } \mu \ll \mu^* \\ +\infty & \text{otw.} \end{cases}$

Then (formally)

$$H(\mu|\mu^*) = \int \left(\log \frac{d\mu}{dx} - \log \frac{d\mu^*}{dx} \right) d\mu$$

$$= \int \frac{d\mu}{dx} \log \frac{d\mu}{dx} \, dx + \int -\log \frac{d\mu^*}{dx} \, d\mu$$

$$= H[\mu] + V[\mu]$$

for $\Psi := -\log \frac{d\mu^*}{dx}$. $(+\log Z)$

So W_2-GF for $H(\mu|\mu^*)$ is $\boxed{\partial_t \mu_t = \operatorname{div}(\nabla \Psi * \mu_t) + \Delta \mu_t}$

"Fokker-Planck eq$^\circ$" (?)

D.II) W_2 Gradient Descent

Prop: The exponential map is $\mathrm{Exp}_\mu(s) = [id + \nabla(Ls)]_\# \mu$

"Proof": Let $s \in T_\mu \mathcal{M}$, i.e. $\int_{\mathbb{R}^d} s = 0$. Want to find a geodesic $(\mu_t)_t$ s.t $\partial_t \mu_t \big|_{t=0} = s$. Look for it in the form $\mu_t = [\underbrace{(1-t)x + tT(x)}_{=:T_t}]_\# \mu$.

We know that (cf proof of Benamou-Brenier formula) $(\mu_t)_t$ satisfies

$$\partial_t \mu_t = -\mathrm{div}(v_t \mu_t) \quad \text{with} \quad v_t = (T - id) \circ T_t^{-1}.$$

So $(\mu_t)_t$ achieves $\left(\mu_{t=0}, \partial_t \mu_t\big|_{t=0}\right) = \left(\mu, -\mathrm{div}\left((T-id)\mu\right)\right)$

$$\overset{!}{=} (\mu, s)$$

$\leadsto$ Take T s.t $-\mathrm{div}\left((T-id)\mu\right) = s \overset{\text{by def.}}{=} -\mathrm{div}\left(\nabla(Ls)\mu\right)$

$$T - id = \nabla(Ls).$$

Then
$$\mu_t = [(1-t)x + tT(x)]_\# \mu = [id + t(T-id)]_\# \mu$$
$$= [id + t\nabla(Ls)]_\# \mu$$
$$\mu_1 = [id + \nabla(Ls)]_\# \mu.$$

To conclude that $\mathrm{Exp}_\mu(s) = \mu_1$ we still need to check that $(\mu_t)_t$ is indeed a geodesic ... Not hard but requires (AFAICT) talking about Kantorovich potentials. [F+21, Rk 2.5.9] $\square$

Def: W_2-GD for $F[\mu]$ is $\boxed{\mu^{k+1} = \left[id - \nabla\left(F'_{\mu^k}\right)\right]_\# \mu^k}$

Indeed, $s = -\mathrm{grad}_W F[\mu^k] = +\mathrm{div}\left(\nabla\left(F'_{\mu^k}\right)\cdot\mu^k\right) = -L^{-1} F'_{\mu^k}$

$$\Rightarrow \mathrm{Exp}_\mu(s) = \left[id + \nabla(Ls)\right]_\# \mu$$
$$= \left[id - \nabla\left(L L^{-1} F'_\mu\right)\right]_\# \mu$$
$$= \left[id - \nabla\left(F'_\mu\right)\right]_\# \mu .$$

Rk: if $\mu^k = \sum_{i=1}^n a_i \delta_{x_i^k}$, then $\mu^{k+1} = \sum_{i=1}^n a_i \delta_{x_i^{k+1}}$

[and some conditions on F],

with $\boxed{\forall i, \quad x_i^{k+1} = x_i^k - \nabla\left(F'_{\mu^k}\right)(x_i^k) .}$

Rk: The above does not apply for $F[\mu] = H[\mu]$

or $F[\mu] = H(\mu \mid \mu^*)$!

(This is still consistent w/ Riemannian formalism: it's just that $H[\mu]$ does not have a first variation at μ's of the form $\sum_i a_i \delta_{x_i}$.)

Indeed, recall its W_2-GF is just heat eq°: $\partial_t \rho_t = \Delta \rho_t$,

for which (distributionally) $\rho_0 = \delta_{x_0} \Rightarrow \rho_t = \frac{1}{Z_t} e^{-\frac{1}{2t}\|\cdot - x_0\|^2}$.

Instead we need to introduce smoothing ... [Liu et al., 2019: "Understanding & accelerating Particle-based VI"]

Rk : ⊛ We did not prove uniqueness of $\mathrm{Exp}_\mu(s)$ ...

⊛ Some geodesics are never part of an exponential map of the form described above, e.g. $\mu_t = (\Sigma_t)_\# \pi^*$ with $\{\pi^*\} = \Gamma_{opt}(\mu_0, \mu_1)$

⊛ Question to Riemannian geometry experts : what is causing this weirdness ?

~~→ did we perhaps "fail to describe all of" $T_\mu \mathcal{M}$ and $T_\mu^* \mathcal{M}$?~~

⊛ Related weirdness : when $\mu = \sum_{i=1}^{n} \delta_{x_i}$, $T_\mu \mathcal{M}$ morally has dimension n, since n is the nb of degrees of freedom in the choice of the geodesics $(T_t)_\# \mu$ ...

D.III) Displacement convexity

Displacement convexity = geodesic convexity in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$.

<u>Prop</u>: • $H[\mu] =$ « $\int_{\mathbb{R}^d} \mu \log \mu$ » is g. convex

• $V[\mu] = \int_{\mathbb{R}^d} \Psi \, d\mu$ is g. λ-convex iff Ψ is λ-convex

• $H(\mu \,|\, \mu^*) = H[\mu] + V[\mu]$ is g. λ-convex if $\log \frac{d\mu^*}{dx}$ is λ-concave

<u>Rk</u>: The most commonly considered functionals on $\mathcal{P}_2(\mathbb{R}^d)$ are

→ $H[\mu]$ entropy, $V[\mu]$ potential energy already discussed

→ Interaction energy functional:

$$W[\mu] = \iint k(x - x') \, d\mu(x) \, d\mu(x')$$

→ Internal energy functional:

$$U[\mu] = \begin{cases} \int \phi\left(\frac{d\mu}{dx}\right) dx & \text{if } \mu \ll \text{Leb.} \\ +\infty & \text{otw.} \end{cases}$$

The criteria for g. convexity of $W[\mu]$ and $U[\mu]$ are known. [F+21, §4.3] [A+21, Lec. 15]

D. IV) W_2-PL for relative entropy is log-Sobolev

TODO

Further topics ...

- will be able to talk abt this
- More functional inequalities
- can talk about this
- Particle discretizations of W_2-GF and W_2-GD
- can talk about this
- Unbalanced OT, Wasserstein-Fisher-Rao distance
- don't know anything
- Entropy-regularized OT & Schrödinger bridge

[A+21 Lec. 19]

- don't know anything
- OT on manifolds, heat flow, connection to Ricci curvature