W_2 gradient flows and PDE

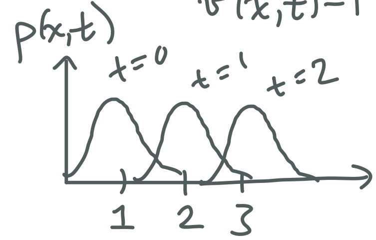
$d\rho(x) = \rho(x)\,dx$

Motivation: $\rho(x,t) \geq 0$, $\int \rho(x,t)\,dx = 1$

$x_1(t), \dots, x_n(t) \in \mathbb{R}^d$

⓪ Continuity equation

$v(x,t) = 1$



PDE: $\begin{cases} \partial_t \rho + \nabla \cdot (v\rho) = 0 \\ \rho(x,0) = \rho_0(x) \end{cases}$

$t \in [0,T]$

Particles: $\begin{cases} \dot{x}_i(t) = v(x_i(t), t) \\ x_i(0) = x_{i,0} \end{cases}$ $\quad W_2(\rho^N(t), \rho(t)) \leq C_{T, \|\nabla v\|_\infty} W_2(\rho^N(0), \rho(0))$

$$\boxed{\begin{array}{l} \text{For } v \text{ nice, if } \rho^N(0) = \frac{1}{N}\sum_{i=1}^N \delta_{x_{i,0}} \xrightarrow{N \to +\infty} \rho_0(x), \\ \text{then } \rho^N(t) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i(t)} \xrightarrow{N \to +\infty} \rho(x,t). \end{array}}$$

① Aggregation eqn w/ drift [AGS'05, CDFLS'10]

$v(x)$ $\quad W(x) = \pm \frac{|x|^{2-d}}{2-d}$, $\frac{|x|^a}{a} - \frac{x^b}{b}$, $-d < b \leq a$

PDE: $\partial_t \rho = \nabla \cdot (\nabla V \rho) + \nabla \cdot ((\nabla W * \rho)\rho)$

Particles: $\dot{x}_i(t) = -\nabla V(x_i(t)) - \frac{1}{N}\sum_{j=1}^N \nabla W(x_i(t) - x_j(t))$

Energy: $E(\rho) = \int V(x)\rho(x)\,dx + \frac{1}{2}\iint W(x-y)\rho(x)\rho(y)\,dx\,dy$

② Fokker–Planck eqn:

$\partial_t \rho + \nabla\cdot(\underbrace{(-\nabla V) + (-\tfrac{\nabla \rho}{\rho})}_{v(x,t)}\rho) = 0$

PDE: $\partial_t \rho = \nabla\cdot(\nabla V \rho) + \Delta \rho$

Particles: $dX_t = \sqrt{2}\, dB_t - \nabla V(X_t)\,dt$

Energy: $E(\rho) = \int V(x)\rho(x)\,dx + \int \rho(x)\log\rho(x)\,dx$

③ 2-layer NN: [MMN'18], [RVE'18], [CB'18], ..

$\Phi(x,z) = x_i\left(\sum_i x_i z_i + x_d\right)_+$

$\Phi(x,z) = \psi(|x-z|)$

Energy: $E(\rho) = \frac{1}{2}\int \left|\int \Phi(x,z)\rho(x)\,dx - f_0(z)\right|^2 d\nu(z)$

$$= \frac{1}{2}\iint \overbrace{\int \Phi(x,z)\Phi(y,z)\,d\nu(z)}^{\omega(x,y)}\, \rho(x)\rho(y)\,dx\,dy$$

$$- \int \underbrace{\int \Phi(x,z) f_0(z)\,d\nu(z)}_{V(x)}\, \rho(x)\,dx + C$$

④ Blob method for χ^2 GF [CLLMR, '20], [CEHT, in prep]

Energy: $E(\rho) = \frac{1}{2}\int \frac{|\rho(x) - \bar\rho(x)|^2}{\bar\rho(x)}\,dx$

Another energy: $E_\varepsilon(\rho) = \frac{1}{2}\int \frac{|\varphi_\varepsilon * \rho(x) - \bar\rho(x)|^2}{\bar\rho(x)}\,dx$

Moral: W_2 perspective brings new tools that complement pure PDE/particle pespective

---

Plan

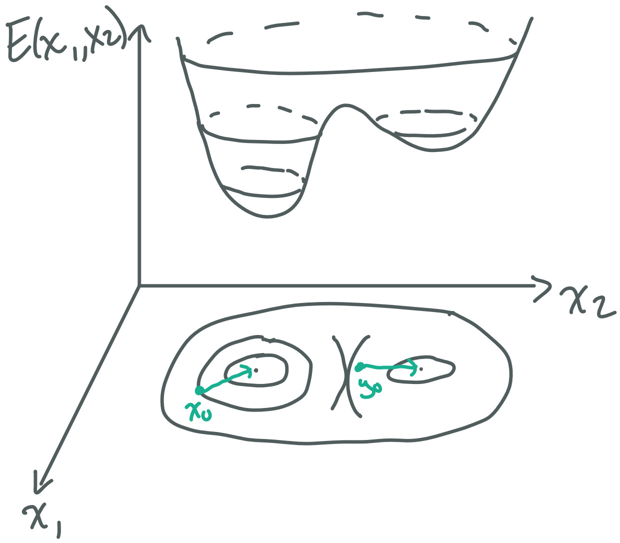
1) Warmup: GF on $\mathbb{R}^d$
2) GF and PDE
3) GF on metric spaces
4) Wasserstein GF

Warmup: GF on $\mathbb{R}^d$

Consider $E: \mathbb{R}^d \to \mathbb{R}$, $x_0 \in \mathbb{R}^d$.

Def: A gradient flow of E with initial conditions x_0 is a solution of

$$\begin{cases} \dot{x}(t) = -\nabla E(x(t)), & t \geq 0 \\ x(0) = x_0 \end{cases}$$



EXISTENCE

- $E \in C^1(\mathbb{R}^d)$ (at least for short time)

UNIQUENESS/ $t \to +\infty$

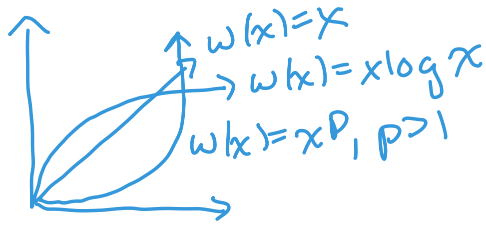
- one sided Lipschitz: $\exists\ \lambda \in \mathbb{R}$ s.t.

$$\langle x-y, \nabla E(x) - \nabla E(y) \rangle \geq \lambda |x-y|^2, \quad \forall\, x,y \in \mathbb{R}^d$$

$E \in C^1$

- above tangent line ineq: $\exists\, \lambda \in \mathbb{R}$ s.t.
  $$E(y) - E(x) - \langle \nabla E(x), y-x \rangle \geq \frac{\lambda}{2}|x-y|^2 \quad \forall x,y \in \mathbb{R}^d$$
  for $\omega(x)$ cts, increasing, $\omega(0)=0$ $\qquad \frac{\lambda}{2}\omega(|x-y|)^2$

- λ-convex: $\exists\, \lambda \in \mathbb{R}$ s.t.
  $$E((1-\alpha)x + \alpha y) \leq (1-\alpha)E(x) + \alpha E(y) - (1-\alpha)\alpha \frac{\lambda}{2}|x-y|^2 \quad \forall x,y \in \mathbb{R}^d,\ \alpha \in [0,1]$$



Why does this imply uniqueness?

Suppose $x(t)$ and $y(t)$ are gradient flows of E.

$$\frac{d}{dt}|x(t)-y(t)|^2 = 2\langle x(t)-y(t), -\nabla E(x(t)) + \nabla E(y(t))\rangle$$
$$\leq -2\lambda |x(t)-y(t)|^2$$

$$|x(t)-y(t)|^{\cancel{2}} \leq e^{-\cancel{2}\lambda t}|x(0)-y(0)|^{\cancel{2}}, \quad t \geq 0$$

$$\frac{d}{dt}a(t) = -2\lambda a(t)$$
$$a(t) = a(0)e^{-2\lambda t}$$

$w(x) = x\log x \qquad \text{RHS} = |x(0)-y(0)|^{e^{2\lambda t}}$

$w(x) = x^p,\ p>1 \qquad \text{RHS} = \left((p-1)2\lambda t + |x(0)-y(0)|^{2(1-p)}\right)^{\frac{1}{2(1-p)}}$

Rmk:

- $\overset{E\in C^2}{\Longleftrightarrow} D^2E \geq \lambda I_{d\times d}$
- E_0 is λ_0-convex, E_1 is λ_1-convex, then E_0+E_1 is $(\lambda_0+\lambda_1)$-convex
- $\lambda>0 \Rightarrow E$ is strongly convex
- $(1-\alpha)x+\alpha y$ is the geodesic from x to y

GENERALIZE TO METRIC SPACE?

Energy dissipation equality (EDE)

$$\dot{x}(t) = -\nabla E(x(t)) \Longleftrightarrow \begin{array}{l} |\dot{x}(t)| = |\nabla E(x(t))| \\ \frac{d}{dt}E(x(t)) = -|\nabla E(x(t))|\,|\dot{x}(t)| \end{array}$$

$ab \geq -\frac{a^2}{2} - \frac{b^2}{2}$ with "=" only if $a=-b$

$$\Longleftrightarrow \frac{d}{dt}E(x(t)) \leq -\frac{1}{2}|\nabla E(x(t))|^2 - \frac{1}{2}|\dot{x}(t)|^2$$

Time discretization

Suppose E lower semicts, λ-convex.

$$\begin{cases} \dot{x}(t) = -\nabla E(x(t)) \\ x(0) = x_0 \end{cases} \xrightarrow[\text{Euler}]{\text{Implicit}} \begin{array}{l} \frac{x_n - x_{n-1}}{\tau} = -\nabla E(x_n) \quad n\geq 1 \\ x_0 = x_0 \end{array}$$

Given x_{n-1},

$$x_n = \underset{x}{\operatorname{argmin}} \left\{ \underbrace{\overbrace{\frac{1}{2\tau}|x - x_{n-1}|^2}^{\frac{1}{\tau}\text{ convex}} + \overbrace{E(x)}^{\lambda\text{convex}}}_{\frac{1}{\tau}+\lambda \text{ convex}} \right\}$$ "Proximal map"

- energy dissipation: $$\frac{E(x_n) - E(x_{n-1})}{\tau} \leq \frac{-1}{2\tau^2} |x_n - x_{n-1}|^2$$

$$\xrightarrow{\tau = \frac{t}{n},\ n \to +\infty} e^{-\lambda t}$$

- contraction: $$|x_n - y_n| \leq \left(\frac{1}{1+\lambda\tau}\right)^{n-1} |x_0 - y_0|$$

GF and PDE

Ex:

metric	$(L^2(\mathbb{R}^d), \|\cdot\|_{L^2})$	$(\mathcal{P}_2(\mathbb{R}^d), W_2)$
gradient	$\nabla_{L^2} E(f) = \frac{\delta E}{\delta f}$	$\nabla_{W_2} E(\rho) = -\nabla\cdot\left(\rho \nabla \frac{\delta E}{\delta \rho}\right)$
energy	$E(f) = \frac{1}{2}\int \|\nabla f\|^2$	$E(\rho) = \frac{1}{m-1}\int \rho^m + \int V\rho + \frac{1}{2}\int (W * \rho)\rho$ $m=1 \rightarrow \int \rho\log\rho$
GF	$\partial_t f = \Delta f$	$\partial_t \rho = \Delta \rho^m + \nabla\cdot((\nabla V + \nabla W * \rho)\rho)$

GF structure helps with $E(\rho) = \int V\rho$

GF on metric spaces

(X, d) complete metric space

What is analogue of $\dot{x}(t) = -\nabla E(x(t))$?

$$\Updownarrow \ \mathbb{R}^d$$

$$\frac{d}{dt} E(x(t)) \leq -\frac{1}{2}|\nabla E(x(t))|^2 - \frac{1}{2}|\dot{x}(t)|^2$$

Consider $x : [0,T] \to X$, $E : X \to \mathbb{R} \cup \{+\infty\}$

- metric derivative: $|x'|(t) = \lim_{s \to t} \frac{d(x(s), x(t))}{|s-t|}$

  $s_+ = \max\{0, s\}$

- metric slope: $|\partial E|(x) = \limsup_{y \to x} \frac{(E(x) - E(y))_+}{d(x,y)}$
- $x \in AC^2([0,T], X)$ in case $|x'|(t) \in L^2([0,T])$.

Def: $x \in AC^2([0,T], X)$ is the curve of maximal slope of E with initial data x_0 if $x(0) = x_0$ and

$$E(x(t)) - E(x(s)) \leq -\frac{1}{2}\int_s^t |\partial E|^2(x(r))\,dr - \frac{1}{2}\int_s^t |x'|^2(r)\,dr$$

energy decreasing

$\forall\ 0 < s \leq t \leq T$

Questions

- existence?
- uniqueness?
- key properties?
- coincide w/ W_2-GF for $(X,d) = (\mathcal{P}_2(\mathbb{R}^d), W_2)$?

Def: $x:[0,1] \to X$ is a *geodesic* if
$d(x(\alpha), x(\beta)) = |\alpha - \beta| \, d(x(0), x(1))$.

Def: $E: X \to \mathbb{R} \cup \{+\infty\}$ is *λ-convex* if $\forall\, x_0, x_1 \in X$ $\exists$ geodesic $x(\alpha)$ from x_0 to x_1 s.t. $\forall\, \alpha \in [0,1]$,

$$E(x(\alpha)) \leq (1-\alpha) E(x(0)) + \alpha E(x(1)) - (1-\alpha)\alpha \frac{\lambda}{2} d(x_0, x_1)^2$$

$\boxed{\text{TIME DISCRETIZATION}}$

Given x_{n-1},

$$x_n = \underset{x}{\operatorname{argmin}} \Big\{ \underbrace{\frac{1}{2\tau} d(x, x_{n-1})^2 + E(x)}_{\Phi_{x_{n-1}}(x)} \Big\}$$

"Minimizing Movements" [DG, '93], "JKO" [JKO, '98]

Well-defined?

Non-positive curvature metric space

- if $x \mapsto \frac{1}{2} d(x, x_0)^2$ is 1-convex $\forall\, x_0 \in X$, then E λ-convex is sufficient.
- what about if $x \mapsto \frac{1}{2} d(x, x_0)^2$ is nonconvex?

it is λ-convex and

$\swarrow$ λ-cgg

<u>Def</u>: $E: X \to \mathbb{R} \cup \{+\infty\}$ is <u>λ-convex along generalized geodesics</u> if $\forall x_0, x_1, \tilde{x}$, $\exists$ a curve $x(t)$ from x_0 to x_1 s.t. $\Phi_{\tilde{x}}(x)$ is $(\frac{1}{\tau}+\lambda)$-convex for $\frac{1}{\tau}+\lambda > 0$.

$D(E) = \{x : E(x) < +\infty\}$

<u>Thm</u> <u>(AGS'05)</u> If E is lsc and λ-cgg, $\forall x_0 \in D(E)$, $\exists !$ curve of max slope. Also

- EDE, $\frac{d}{dt} E(x(t))$ exists
- $d(x(t), y(t)) \leq e^{-\lambda t} d(x(0), y(0))$
- time discretization converges $O(\tau)$.

Wasserstein GF

$(\mathcal{P}_2(\mathbb{R}^d), W_2)$

How does a curve of max slope relate to PDE?

Three key properties of W_2: [AGS'05]

① if $\rho \in AC^2([0,1], \mathcal{P}_2(\mathbb{R}^d))$, $\exists\, v(x,t)$ s.t. $\partial_t \rho + \nabla \cdot (v \rho) = 0$ and $\left(\int |v(x,t)|^2 \rho(x,t)\, dx\right)^{1/2} = |\rho'|(t)$

PDE Benamou-Brenier for a.e. t

Suppose E is lsc and λ-cgg.

pseudo-Riemannian structure gives chain rule.

(2) For $\xi \in \partial E(\rho(t))$, $\frac{d}{dt} E(\rho(t)) = \int \langle \xi(x), v(x,t) \rangle \rho(x,t)\, dx$

Suppose $\frac{\delta E}{\delta \rho}$ is "well-defined".

(3) $\nabla \frac{\delta E}{\delta \rho} \in \partial E(\rho)$ and $|\partial E|(\rho) = \left(\int |\nabla \frac{\delta E}{\delta \rho}|^2 d\rho \right)^{1/2}$

Nice expression

Then...

Thm [AGS'05] If $\rho(x,t)$ is the curve of max slope of E, then $\partial_t \rho - \nabla \cdot (\nabla \frac{\delta E}{\delta \rho} \rho) = 0$.

Pf: By (1), there exists $v(x,t)$ s.t. $\partial_t \rho + \nabla \cdot (v\rho) = 0$. Thus it suffices to show $v = -\nabla \frac{\delta E}{\delta \rho}$.

$$\frac{d}{dt} E(\rho(t)) = \int \langle \underbrace{\xi(x)}_{\overset{(3)}{=}\nabla \frac{\delta E}{\delta \rho}}, v(x,t) \rangle \rho(x,t)\, dx$$

$ab \geq -\frac{a^2}{2} - \frac{b^2}{2}$ with "=" only if $a = -b$

$$\geq -\frac{1}{2} \|\nabla \frac{\delta E}{\delta \rho}\|^2_{L^2(\rho)} - \frac{1}{2} \|v\|^2_{L^2(\rho)}$$

$$= -\frac{1}{2} |\partial E|^2(\rho(t)) - |\rho'|^2(t)$$

$$\geq \frac{d}{dt} E(\rho(t)).$$

Thus $v = -\nabla \frac{\delta E}{\delta \rho}$, ρ-a.e. $\square$

References

1) Villani, Topics in Optimal Transportation.

2) Ambrosio, Gigli, Savaré, Gradient Flows in Metric Spaces and the Space of Probability Measures

3) Ambrosio, Savaré, Gradient Flows of Probability Measures

4) Santambrogio, Optimal Transport for Applied Mathematicians

5) Santambrogio, Euclidean, Metric, and Wasserstein Gradient Flows

6) Peyré and Cuturi, Computational Optimal Transport

7) Figalli Glaudo, An Invitation to Optimal Transport, W-Distances, and GFs

8) Brezis, Functional Analysis, Sobolev Spaces and PDEs