Gradient Flows in Wasserstein Space and PDEs

SPAM Lab – 2019·09·30

---

* Gradient Flows:

$\dot{x} = -\nabla F(x)$ w/ $x(0) = x_0 \in \mathbb{R}^n$.

Autonomous ODE $\Rightarrow \exists!$ soln.

Euler's Method:

$$x^{\varepsilon}_{n+1} = x^{\varepsilon}_n - \varepsilon \nabla F(x^{\varepsilon}_n)$$

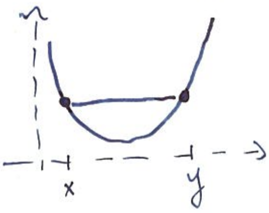
or implicit Euler: $x^{\varepsilon}_{n+1} = x^{\varepsilon}_n - \nabla F(x^{\varepsilon}_{n+1})$ approximate solns.

If ε varies, this is gradient descent for minimizing F.

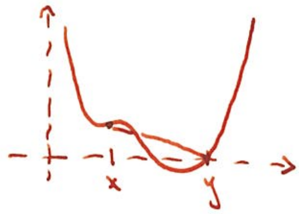
Goal: Tell same story for PDEs.

* Convex Functions:

$$F((1-\tau)x + \tau y) \le (1-\tau)F(x) + \tau F(y) - \frac{\lambda}{2}\tau(1-\tau)\|x-y\|_2^2$$



└ Secant $\ge$ graph



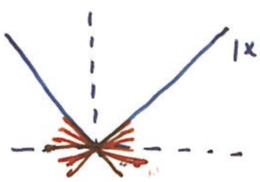
λ-convex w/ $\lambda < 0$.

Secant $+ O(x-y)^2 \ge$ graph.

Replace ∇ with subdifferential: $v \in \partial F(x)$ iff

$$F(x) + \langle v, y-x\rangle \le F(y) - \frac{\lambda}{2}\|x-y\|_2^2 \quad \text{for all } y.$$

E.g. $F(x) = |x|$ on $\mathbb{R}$. Then $\partial F(0) = [-1, 1]$



(Any line w/ slope $-1 \le m \le 1$ is "tangent" at $x=0$.)

Note: if $\nabla F(x)$ exists, $\nabla F(x) = \partial F(x)$.

Gradient flow: $\dot{x}(t) \in -\partial F(x)$ a.e. $t > 0$, $\lim_{t \searrow 0} x(t) = x_0$.

Want $x(t)$ to be 2AC: $\exists f$ s.t. $\|x(t) - x(s)\|_2 \le \int_t^s f(r)\,dr \quad \forall\, t < s$.

* Wasserstein Space:

$$\mathcal{M} = \left\{ \text{prob. measures } \mu \text{ on } \mathbb{R}^n \;\middle|\; \int_{\mathbb{R}^n} \|x\|_2^2 \, d\mu < +\infty \right\} \quad \text{(Finite 2nd-moment)}$$

Wasserstein metric W_2:

$$W_2^2(\mu,\nu) := \inf_{\gamma} \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x-y\|_2^2 \, d\gamma(x,y)$$

where $\pi^1_{\#}\gamma = \mu$ and $\pi^2_{\#}\gamma = \nu$

i.e. marginals of γ are μ & ν.

γ an "optimal coupling" of μ and ν. Exists by Kantorovich duality.

* Structure of $\mathcal{M}$:

Not a vector space! (Obvious, but Riesz-Rep condns otherwise)

Is a geodesic space. There is a "straight-line" between μ, ν given by

$$\alpha(t) = \left((1-t)\pi^1 + t\pi^2\right)_{\#}\gamma_{opt}$$

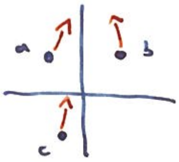
Is not non-positively curved! (In sense of Alexandrov)

Derivatives of curves $x(t): I \to \mathcal{M}$ by

$$|\dot{\mu}|(t) = \lim_{h\to 0} \frac{W_2(\mu_t, \mu_{t+h})}{|h|} \quad \text{(Exists a.e. if } \mu(t) \text{ is 2AC)}$$

Tangent space $T_\mu \mathcal{M} := \{\nabla\varphi \mid \varphi \in C_c^\infty(\mathbb{R}^n)\}$ (then take closure in $L^2(\mathbb{R}^n, \mathbb{R}^n; \mu)$)

(Vector fields on $\mathbb{R}^n$ integrable against μ)



$\mu = \frac{1}{3}\delta_a + \frac{1}{3}\delta_b + \frac{1}{3}\delta_c$

$T_\mu \mathcal{M} = \bigoplus_{i=0}^{3} \mathbb{R}^n$

Hilbert space w/ $\langle f, g \rangle_\mu := \int_{\mathbb{R}^n} \langle f, g \rangle \, d\mu$

$\mathcal{M}$ is like a Riemannian manifold.

* Gradient Flows in $\mathcal{M}$:

$|\dot{\mu}(t)| \in -\partial F(\mu)$ where F is λ-geodesically convex:

$$F(\mu(t)) \le (1-t)F(\mu_0) + tF(\mu_1) - \frac{\lambda}{2}t(1-t)W_2^2(\mu_0, \mu_1) \quad \text{For any geodesic } \mu.$$

and $\partial F(\mu)$ are v.f.s $\underline{V}$ s.t.

$$F(\mu) + \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle \underline{V}(x), y-x \rangle \, d\gamma_{opt} \le F(\nu) - \frac{\lambda}{2} W_2^2(\mu, \nu) \quad \forall \nu \in \mathcal{M}$$

Thm: Gradient flows exist, and are unique, for λ-geo. convex F.

Pf: Define a discrete soln by $\mu_{n+1}^\varepsilon \in \operatorname{argmin}_\mu \left\{ \mu \mapsto F(\mu) + \frac{W_2^2(\mu, \mu_n^\varepsilon)}{2\varepsilon} \right\}$

and show possible to pass to limit $\varepsilon \to 0$. (compare w/ implicit Euler) $\blacksquare$

(Uniqueness ~ technical)

* Application to PDEs:

Define $E: \mathcal{M} \to \mathbb{R}$ by $E(\mu) = \int \rho \log \rho \, d\lambda$ if $\mu = \rho \, d\lambda$ and $+\infty$ otherwise.

Lemma: E is convex, and $\partial E = \nabla \rho$.

Thm: Gradient characterised by $\frac{d}{dt}\mu_t + \nabla \cdot (\underline{V}_t \mu_t) = 0$

Cor: $\dot{\rho} = -\nabla \cdot \nabla \rho$

Three classic functionals:

$$\mathcal{V}(\mu) := \int V(x)\,d\mu \quad \text{for some } V:\mathbb{R}^n \to \mathbb{R}$$

$$\mathcal{W}(\mu) := \iint W(x-y)\,d\mu\otimes d\mu(x,y) \quad W:\mathbb{R}^n \to \mathbb{R}$$

$$\mathcal{E}(\mu) := \int E(\rho)\,dx \qquad \mu \ll \lambda,\ \mu = \rho\lambda.$$

$$+\infty \quad \text{if } \mu \not\ll \lambda.$$

These inherit convexity from their integrands. They have gradients v_t characterized by $\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0$ along gradient flow μ_t

Specifically:

$$\partial\mathcal{V}(\mu) = \nabla V$$

$$\partial\mathcal{W}(\mu) = (\nabla W) * \mu$$

$$\partial\mathcal{E}(\mu) = \nabla(E'(\rho))$$

Eg. If $E(x) = x\log(x)$, then $\partial\mathcal{E}(\mu) = \nabla(\log(\rho)+1) = \nabla\rho$ so the gradient flow of $\mathcal{E}$ is $\frac{d}{dt}\mu_t + \nabla\cdot(\nabla\mathcal{E}\mu_t) = \frac{\partial\rho}{\partial t} + \nabla\cdot(\nabla\rho) = 0$

i.e. the heat equation.