Non-interacting particle systems and Wasserstein gradient flows

Christian Léonard

Université Paris Nanterre

Sylvie 60
IHP - September, 2018

Aim of the talk

the heat flow is the Wasserstein gradient flow of the entropy

- gradient flow
- Wasserstein
- convexity inequalities
- large deviations do a good job

References

- Jordan-Kinderlehrer-Otto, Otto, Otto-Villani, von Renesse-Sturm
- Ambrosio-Gigli-Savaré
- Adams-Dirr-Peletier-Zimmer
- joint work with J. Zimmer

Gradient flow in $\mathcal{X}$

- $\mathcal{X} = \mathbb{R}^n$

gradient flow equation

$$\begin{cases} \dot{\theta}_t &= -F'(\theta_t), \quad t \geq 0 \\ \theta_0 &= x_o \end{cases}$$

- hypothesis: F is K-convex $\iff F'' \geq K\mathrm{Id}, \quad K \in \mathbb{R}$
- semigroup: $\theta_t =: S_t(x_o), \quad t \geq 0$

- Lyapunov function: $\frac{d}{dt}F(\theta_t) = -I(\theta_t) \leq 0, \qquad I := |F'|^2$
- contraction: $|S_t(y) - S_t(x)| \leq e^{-Kt}|y - x|, \quad \forall t, x, y$
- relaxation to equilibrium: $|S_t(x) - x_*| \leq e^{-Kt}|x - x_*|, \quad \forall t, x$
  - $K > 0$
  - $F'(x_*) = 0$

Gradient flow in $\mathcal{X}$

$$\begin{aligned}
& F \text{ is } K\text{-convex} \\
\iff \quad & t \in [0,1] \mapsto F(\gamma_t^{xy}) \quad \text{is } K|y-x|^2\text{-convex}, \quad \forall x, y \\
\iff \quad & F'(x)\cdot(y-x) + \frac{K}{2}|y-x|^2 \le F(y) - F(x), \quad \forall x, y \\
\iff \quad & \frac{d}{dt}_{|t=0} \frac{1}{2}|y - S_t(x)|^2 + \frac{K}{2}|y-x|^2 \le F(y) - F(x), \quad \forall x, y
\end{aligned}$$

- geodesic: γ^{xy}
- recall: $\frac{d}{dt}_{|t=0} S_t(x) = -F'(x), \quad I := |F'|^2$

$$F(x) - F(y) \le \sqrt{I(x)}|y-x| - K|y-x|^2/2, \qquad \forall x, y$$

Gradient flow in $\mathcal{X}$

$$F(x) - F(y) \le \sqrt{I(x)}|y - x| - K|y - x|^2/2, \qquad \forall x, y$$

- equilibrium: $F'(x_*) = 0,$ normalization: $F(x_*) = 0$

- $K|y - x_*|^2/2 \le F(y), \quad \forall y$ (Talagrand)
- $F(x) \le \sqrt{I(x)}|x - x_*| - K|x - x_*|^2/2, \quad \forall x$ (HWI)
- $F(x) \le I(x)/(2K), \quad \forall x$ (log-Sobolev)

- relaxation to equilibrium: $|S_t(x) - x_*| \le e^{-Kt}|x - x_*|, \quad \forall t, x$
- $t \mapsto F(S_t(x))$ decreases

$\mathbb{R}^n \rightsquigarrow \mathrm{P}(\mathcal{X})$

- F: free energy, I: Fisher information
- $|\cdot| \rightsquigarrow W_2$: Wasserstein distance
- $(S_t)_{t\ge0}$: Fokker-Planck semigroup
  - S is the W_2-gradient flow of F

Gradient flow in $\mathcal{X}$

De Giorgi minimizing movement

- time step: $\tau > 0$
- $\xi^\tau_{k+1} = \operatorname{argmin} j^\tau(\xi^\tau_k, \cdot), \quad k \geq 0$
  - $j^\tau(x, y) := F(y) - F(x) + \tau^{-1}|y - x|^2/2$
- $\theta^\tau_t := \xi^\tau_{\tau\lfloor t/\tau\rfloor}, \quad t \geq 0, \qquad \theta^\tau_0 = \xi^\tau_0 = x_o$

$$\lim_{\tau\downarrow 0} \theta^\tau_t = S_t(x_o), \quad t \geq 0$$

- hint: $\partial_y j^\tau(x, y) = 0 \iff (y - x)/\tau = -F'(y)$

Large deviations in $\mathcal{C}_\mathcal{X}$

- $\mathcal{C}_\mathcal{X} := C([0,\infty),\mathcal{X})$

$$\begin{cases} d\mathrm{X}^\epsilon_t = -F'(\mathrm{X}^\epsilon_t)\,dt + \sqrt{\epsilon}\;d\mathrm{W}_t, & t \geq 0, \\ \mathrm{X}^\epsilon_0 = x_o \end{cases}$$

- almost sure limit: $\lim_{\epsilon\to 0} \mathrm{X}^\epsilon = S(x_o)$

$$\mathbb{P}(\mathrm{X}^\epsilon \in A) \underset{\epsilon\to 0}{\asymp} \exp\left(-\epsilon^{-1} \inf_A I\right), \quad A \subset \mathcal{C}_\mathcal{X}$$

- $I(\omega) = \iota_{\{\omega_0 = x_o\}} + \frac{1}{2}\int_{[0,\infty)} |\dot{\omega}_t + F'(\omega_t)|^2 \; dt$
- $\operatorname{argmin} I = S(x_o)$

Large deviations in $\mathcal{C}_{\mathcal{X}}$

- $\mathcal{X} = \mathbb{R}^n$
- time discretization: $\quad \mathrm{X}^{\epsilon,\tau}_t := \mathrm{X}^{\epsilon}_{\lfloor t/\tau \rfloor \tau}, \quad t \geq 0$

$$\mathbb{P}(\mathrm{X}^{\epsilon,\tau} \in \bullet) \underset{\epsilon \to 0}{\asymp} \exp\left(-\epsilon^{-1} \inf_{\bullet} I^\tau \right)$$

- $I^\tau(\eta) = \inf \{ I(\tilde{\eta});\ \tilde{\eta}^\tau = \eta \}$
  $\qquad = \iota_{\{\eta \in \mathcal{C}^{\tau, x_o}_{\mathcal{X}}\}} + \sum_{k \geq 0} J^\tau(\eta_{k\tau}, \eta_{k\tau+\tau})$
- $J^\tau(x, y)$
  $= \inf \left\{ \frac{1}{2} \int_{[k\tau,(k+1)\tau]} |\dot{\eta}_t + F'(\eta_t)|^2\ dt; \eta : \eta_{k\tau} = x, \eta_{(k+1)\tau} = y \right\}$
  $= F(y) - F(x) + \tau^{-1} C^\tau(x, y)$
- $C^\tau(x, y) = \inf \left\{ \frac{1}{2} \int_{[0,1]} |\dot{\omega}_s|^2\, ds + \frac{\tau^2}{2} \int_{[0,1]} |F'(\omega_s)|^2\, ds\ ;\ \omega \in \Omega^{xy} \right\}$
  - $\Omega^{xy} := \{ \omega \in C([0,1], \mathcal{X}) : \omega_0 = x, \omega_1 = y \}$

Gradient flow in $\mathcal{X}$ revisited

$$C^\tau(x,y) = \inf\left\{ \frac{1}{2}\int_{[0,1]} |\dot\omega_s|^2\,ds + \frac{\tau^2}{2}\int_{[0,1]} |F'(\omega_s)|^2\,ds\ ;\ \omega \in \Omega^{xy} \right\}$$

- $C^{\tau=0}(x,y) = |y-x|^2/2$
- $j^\tau(x,y) = F(y) - F(x) + \tau^{-1} C^{\tau=0}(x,y)$
- $J^\tau(x,y) = F(y) - F(x) + \tau^{-1} C^\tau(x,y)$

LD minimizing movement

- $\xi^\tau_{k+1} = \operatorname{argmin} J^\tau(\xi^\tau_k, \cdot), \quad k \geq 0$
- $\theta^\tau_t := \xi^\tau_{\tau\lfloor t/\tau \rfloor}, \quad t \geq 0, \qquad \theta^\tau_0 = \xi^\tau_0 = x_o$

$$\lim_{\tau\downarrow 0} \theta^\tau_t = S_t(x_o), \quad t \geq 0$$

- proof: $\quad \Gamma\text{-}\lim_{\tau\downarrow 0} I^\tau = I$

Optimal transport

- $\mathcal{X} = \mathbb{R}^n, \quad \mathcal{X} = M$
- $\mu_0, \mu_1 \in \mathrm{P}(\mathcal{X})$
- $\pi \in \Pi(\mu_0, \mu_1) \iff [\pi(dx \times \mathcal{X}) = \mu_0(dx),\ \pi(\mathcal{X} \times dy) = \mu_1(dy)]$

Monge-Kantorovich problem

$$\inf_{\pi \in \Pi(\mu_0,\mu_1)} \int_{\mathcal{X}^2} d(x,y)^2\, \pi(dxdy) =: W_2^2(\mu_0, \mu_1) \qquad (1)$$

Benamou-Brenier formula

- $W_2^2(\mu_0, \mu_1) = \inf_{(\mu, \nabla\phi)} \int_{[0,1]\times\mathcal{X}} |\nabla \phi_t(x)|^2\, \mu_t(dx)dt$
  - $\mu := (\mu_t)_{0 \le t \le 1}, \quad \mu_0, \mu_1$ fixed
  - $\partial_t \mu + \nabla \cdot (\mu \nabla \phi) = 0$

Optimal transport

- $\Omega = C([0,1],\mathcal{X}), \quad P \in \mathrm{P}(\Omega)$

$$\inf_{P:P_0=\mu_0,P_1=\mu_1} E_P \int_{[0,1]} |\dot{X}_t|^2\,dt = W_2^2(\mu_0,\mu_1) \qquad (2)$$

- $E_P \int_{[0,1]\times\mathcal{X}} |\dot{X}_t|^2\,dt = \int_{\mathcal{X}^2} E_{P^{xy}}\left(\int_{[0,1]\times\mathcal{X}} |\dot{X}_t|^2\,dt\right) P_{01}(dxdy)$
  $\geq \int_{\mathcal{X}^2} d(x,y)^2\, P_{01}(dxdy)$
- $P \in \mathrm{sol}(2) \iff \begin{cases} P(\cdot) = \int_{\mathcal{X}^2} \delta_{\gamma^{xy}}(\cdot)\,\pi(dxdy) \\ \pi \in \mathrm{sol}(1) \end{cases}$

$$\iff \begin{cases} \ddot{X}_t = 0, \quad P\text{-a.s.} \\ P_{01} \in \mathrm{sol}(1) \end{cases}$$

- $\mu_t := P_t,\ 0 \leq t \leq 1$ defines a *displacement interpolation*
  - natural candidate for a geodesic in $\mathrm{P}(\mathcal{X})$

Optimal transport

- $P \in \mathrm{sol}(2), \quad \mu_t := P_t$
- P is Markov (uneasy)
  - hint: $v_t(X_t) = E_P(\dot{X}_t \mid X_t)$
  - $dX_t = v_t(X_t)\, dt, \quad P\text{-a.s.}$
  - $\partial_t \mu + \nabla \cdot (\mu v) = 0$
  - no shock
- $v_t \in \overline{\{\nabla \phi\}}^{L^2(\mu_t)}$
  - $dX_t = \nabla \phi_t(X_t)\, dt, \quad P\text{-a.s.}$
  - $\partial_t \mu + \nabla \cdot (\mu \nabla \phi) = 0$

Otto calculus

- $W_2^2(\mu_0, \mu_1) = \inf_{(\mu, \nabla\phi)} \int_{[0,1]\times\mathcal{X}} |\nabla\phi_t(x)|^2 \, \mu_t(dx)dt$
  - $\mu := (\mu_t)_{0\le t\le 1}, \quad \mu_0, \mu_1$ fixed
  - $\partial_t \mu + \nabla \cdot (\mu \nabla\phi) = 0$

- $d(x,y)^2 = \inf\{\int_{[0,1]} |\dot\omega_t|^2_{\omega_t} \, dt; \ \omega \in \Omega : \ \omega_0 = x, \omega_1 = y\}$
  - $\dot\omega_t \in T_{\omega_t} M$

Wasserstein metric (Otto)

- $\dot\mu_t := \nabla\phi_t, \qquad \partial_t\mu + \nabla\cdot(\mu\dot\mu) = 0$
- $T_{\mu_t}\mathrm{P}(\mathcal{X}) := \overline{\{\nabla\phi\}}^{L^2(\mu_t)}$
- $\dot\alpha_t = \nabla\phi_t, \ \dot\beta_t = \nabla\psi_t, \qquad \left\langle \dot\alpha_t, \dot\beta_t \right\rangle_{\mu_t} := \int_{\mathcal{X}} \nabla\phi_t \cdot \nabla\psi_t \, d\mu_t$

W-gradient flow in $\mathrm{P}(\mathcal{X})$

the heat flow is the Wasserstein gradient flow of the entropy

- heat equation: $\partial_t \mu = \Delta \mu$
- relative entropy: $H(\mu|m) := \int_{\mathcal{X}} \log(d\mu/dm)\, d\mu$
- entropy: $H(\mu) := H(\mu|\mathrm{vol}) = \int_{\mathcal{X}} \mu \log \mu \, d\mathrm{vol}$
- gradient flow: $\dot{\mu}_t = -\operatorname{grad}_W H(\mu_t)$
- proof:
  - $\operatorname{grad}_W H(\mu) = \nabla \log \mu$
    - $\frac{d}{dt} H(\mu_t) = \int_{\mathcal{X}} \log \mu_t \; \partial_t \mu_t \, d\mathrm{vol} = -\int_{\mathcal{X}} \log \mu_t \; \nabla \cdot (\mu_t \nabla \phi_t) \, d\mathrm{vol}$

      $= \int_{\mathcal{X}} \nabla \log \mu_t \cdot \nabla \phi_t \, d\mu_t = \langle \nabla \log \mu_t, \dot{\mu}_t \rangle_{\mu_t}$
  - $\dot{\mu}_t = -\operatorname{grad}_W H(\mu_t)$
    - $\dot{\mu}_t = -\nabla \log \mu_t$
    - $0 = \partial_t \mu + \nabla \cdot (\mu \dot{\mu}) = \partial_t \mu - \nabla \cdot (\mu \nabla \log \mu) = \partial_t \mu - \Delta \mu$

W-gradient flow in $\mathrm{P}(\mathcal{X})$

Fokker-Planck equation

$$\partial_t \mu - \nabla \cdot (\mu \nabla V/2) = \Delta \mu/2$$

- $\mu_t = e^{tL}\mu_0, \qquad L = [-\nabla V \cdot \nabla + \Delta]/2$
- equilibrium: $\quad m = e^{-V}\,\mathrm{vol}$
- $R \in \mathrm{P}(\mathcal{C}_{\mathcal{X}})$
  - $R_0 = m$
  - $dX_t = -\frac{1}{2}\nabla V(X_t)\,dt + dW_t^R, \quad R\text{-a.s.}$
  - R is m-reversible
- $R_{\mu_0} := \int_{\mathcal{X}} R_x\,\mu_0(dx) \in \mathrm{P}(\mathcal{C}_{\mathcal{X}})$
  - $\mu_t = (X_t)_\# R_{\mu_0}$

W-gradient flow in $\mathrm{P}(\mathcal{X})$

- $F(\alpha) = H(\alpha|m)/2 = [H(\alpha) + \int_{\mathcal{X}} V\, d\alpha]/2$

$$J^{\tau}_{JKO}(\alpha, \beta) = F(\beta) - F(\alpha) + \tau^{-1} W_2^2(\alpha, \beta)/2$$

JKO minimizing movement

- $\xi^{\tau}_{k+1} = \operatorname{argmin} J^{\tau}_{JKO}(\xi^{\tau}_k, \cdot), \quad k \geq 0$
- $\mu^{\tau}_t = \xi^{\tau}_{\tau\lfloor t/\tau \rfloor}, \quad t \geq 0, \qquad \mu^{\tau}_0 = \xi^{\tau}_0 = x_o$

JKO heuristic result

$$\lim_{\tau \downarrow 0} \mu^{\tau} = \mu$$

Large deviations in $\mathcal{C}_{\mathrm{P}(\mathcal{X})}$

- random dynamical particles: $\mathsf{Z}^1, \mathsf{Z}^2, \dots$ iid(R)
- empirical path measure: $\hat{\mathsf{Z}}^N := \frac{1}{N} \sum_{1 \le i \le N} \delta_{\mathsf{Z}^i} \quad \in \mathrm{P}(\mathcal{C}_{\mathcal{X}})$
- $\lim_{N\to\infty} \hat{\mathsf{Z}}^N(0) = \mu_0, \quad \mathbb{P}\text{-a.s.}$

$$\mathbb{P}(\hat{\mathsf{Z}}^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\left(-N \inf_{P \in \bullet} [\iota_{\{P_0 = \mu_0\}} + H(P|R_{\mu_0})] \right)$$

Large deviations in $\mathcal{C}_{\mathrm{P}(\mathcal{X})}$

- $\mathcal{C}_{\mathrm{P}(\mathcal{X})} := C([0,\infty), \mathrm{P}(\mathcal{X}))$
- empirical process: $\quad \mathrm{Z}^N = [t \in [0,\infty) \mapsto \bar{\mathrm{Z}}^N(t)] \quad \in \mathcal{C}_{\mathrm{P}(\mathcal{X})}$

$$\mathbb{P}(\bar{\mathrm{Z}}^N \in \bullet) \underset{N\to\infty}{\asymp} \exp\left(-N \inf_{\bullet} I\right)$$

- $I(\nu) = \iota_{\{\nu_0=\mu_0\}} + \inf\{H(P|R_{\mu_0}); P \in \mathrm{P}(\mathcal{C}_{\mathcal{X}}) : P_t = \nu_t, \ \forall t\}$
- argmin $I = \mu$: $\quad$ solution of FP equation

Large deviations in $\mathcal{C}_{\mathrm{P}(\mathcal{X})}$

- $I(\nu) = \iota_{\{\nu_0=\mu_0\}} + \inf\{H(P|R_{\mu_0}); P \in \mathrm{P}(\mathcal{C}_{\mathcal{X}}) : P_t = \nu_t, \ \forall t\}$
- $\eta \in \mathcal{C}_{\mathcal{X}}, \quad \eta^\tau_t := \eta_{\tau\lfloor t/\tau \rfloor}$

$$\mathbb{P}(\bar{Z}^{N,\tau} \in \bullet) \underset{N\to\infty}{\asymp} \exp\left(-N \inf_{\bullet} I^\tau\right)$$

- $I^\tau(\nu) = \iota_{\mathcal{C}^\tau_{\nu_0}}(\nu) + \sum_{k\geq 0} J^\tau_{LD}(\nu_{k\tau}, \nu_{(k+1)\tau})$

$$J^\tau_{LD}(\alpha,\beta) = \inf\{H(\mathrm{p}|R_{[0,\tau]}); \mathrm{p} \in \mathrm{P}(\Omega^\tau) : \mathrm{p}_0 = \alpha, \mathrm{p}_\tau = \beta\} - H(\alpha|m)$$

  - $\Omega^\tau = C([0,\tau],\mathcal{X}), \quad R_{[0,\tau]} \in \mathrm{P}(\Omega^\tau)$

LD gradient flow in $\mathcal{C}_{\mathrm{P}(\mathcal{X})}$

- $\Gamma\text{-}\lim_{\tau\downarrow 0} I^\tau = I$
- $\mu^\tau := \operatorname{argmin} I^\tau, \quad \mu:$ solution of FP equation

$$\lim_{\tau\downarrow 0} \mu^\tau = \mu$$

LD minimizing movement

- $\xi^\tau_{k+1} = \operatorname{argmin} J^\tau_{LD}(\xi^\tau_k, \cdot), \quad k \geq 0$
- $\mu^\tau_t = \xi^\tau_{\tau\lfloor t/\tau\rfloor}, \quad t \geq 0, \qquad \mu^\tau_0 = \xi^\tau_0 = x_o$

LD gradient flow in $\mathcal{C}_{\mathrm{P}(\mathcal{X})}$

heuristic result (JKO)

the FP flow is the W-gradient flow of the free energy

rigorous result

the FP flow is the LD-gradient flow of the free energy

LD gradient flow in $\mathcal{C}_{\mathrm{P}(\mathcal{X})}$

$$J^{\tau}_{LD}(\alpha,\beta) = \inf\{H(\mathsf{p}|R_{[0,\tau]}); \mathsf{p} \in \mathrm{P}(\Omega^{\tau}) : \mathsf{p}_0 = \alpha, \mathsf{p}_\tau = \beta\} - H(\alpha|m)$$

- Schrödinger problem
- time reversal, $\overrightarrow{v} = v^{\mathrm{cu}} + v^{\mathrm{os}}, \quad v^{\mathrm{os}} = \nabla \log \sqrt{d\mu/d\mathrm{vol}}$
- $\quad H(\mathsf{p}|R_{[0,\tau]}) - H(\mathsf{p}_0|m)$
  $= E_{\mathsf{p}} \int_{[0,\tau]} |\overrightarrow{v}_t|^2/2\, dt$
  $$= [H(\mathsf{p}_\tau|m) - H(\mathsf{p}_0|m)]/2 + \underbrace{E_{\mathsf{p}} \int_{[0,\tau]} |v^{\mathrm{cu}}_t|^2/2\, dt}_{O(\tau^{-1})} + \underbrace{E_{\mathsf{p}} \int_{[0,\tau]} |v^{\mathrm{os}}_t|^2/2\, dt}_{O(\tau)}$$

$$J^{\tau}_{LD}(\alpha,\beta) = F(\beta) - F(\alpha) \ + \ \tau^{-1} W_2^2(\alpha,\beta)/2 \ + \ O(\tau)$$

- $F(\alpha) = H(\alpha|m)/2$
- $J^{\tau}_{JKO}(\alpha,\beta) = F(\beta) - F(\alpha) + \tau^{-1} W_2^2(\alpha,\beta)/2$

Perspectives

- numerical scheme: J^{τ}_{LD} is used in practice to compute J^{τ}_{JKO}
  - speedlight algorithm (Cuturi)
  - Mokaplan team (Benamou, Peyré, Brenier, Carlier, ...)
- gradient flows/curvature lower bound on graphs
- gradient flows related to nonlinear dissipative PDE's (biology?)
  - reaction-diffusion

happy birthday Sylvie