

Méthodes numériques pour les flots gradients dans l'espace de Wasserstein

Guillaume Carlier ^a

Travaux en collaboration avec J.-D. Benamou, M. Cuturi, V. Duval, M. Laborde, Q. Mérigot, L. Nenna, E. Oudet et B. Schmitzer dans le cadre de l'équipe projet MOKAPLAN.
Journées SMAI-MODE 2016, Toulouse.

^aCEREMADE, Université Paris Dauphine et MOKAPLAN (Inria-Dauphine).

Introduction

Essor important de la théorie du transport optimal depuis le début des années 90: Y. Brenier. Notons $\mathcal{P}_2(\mathbf{R}^d)$ l'ensemble des mesures de probabilités sur \mathbf{R}^d de second moment finis, pour ρ_0 et ρ_1 dans $\mathcal{P}_2(\mathbf{R}^d)$, la 2-distance de Wasserstein entre ρ_0 et ρ_1 , $W_2(\rho_0, \rho_1)$ est donnée par

$$W_2^2(\rho_0, \rho_1) = \inf_{\gamma \in \Pi(\rho_0, \rho_1)} \int_{\mathbf{R}^d \times \mathbf{R}^d} |x - y|^2 d\gamma(x, y)$$

où $\Pi(\rho_0, \rho_1)$ est l'ensemble des plans de transport entre ρ_0 et ρ_1 i.e. les probas sur $\mathbf{R}^d \times \mathbf{R}^d$ ayant ρ_0 et ρ_1 pour marges. W_2 est une distance sur $\mathcal{P}_2(\mathbf{R}^d)$, $(\mathcal{P}_2(\mathbf{R}^d), W_2)$ espace de Wasserstein.

Brenier, McCann: si ρ_0 ne charge pas les hypersurfaces Lipschitz il y a une unique solution, qui est caractérisée par $\gamma = (\text{id}, \nabla u)_\# \rho_0$ avec u convexe. Lien avec Monge-Ampère:

$$\det(D^2 u) \rho_1(\nabla u) = \rho_0, \quad u \text{ convexe.}$$

Théorie de la régularité L. Caffarelli (avancées récentes: Figalli, De Philippis). Interpolation (McCann): courbe de mesures $t \in [0, 1] \mapsto \rho_t = ((1 - t) \text{id} + t \nabla u)_\# \rho_0$, c'est la géodésique entre ρ_0 et ρ_1 . Notion de convexité par déplacement.

Formulation dynamique de Benamou-Brenier:

$$W_2^2(\rho_0, \rho_1) = \inf \int_0^1 \int_{\mathbf{R}^d} |v_t(x)|^2 \rho_t(dx)$$

sous les contraintes:

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \quad \rho|_{t=0} = \rho_0, \quad \rho|_{t=1} = \rho_1.$$

F. Otto: structure formelle de variété Riemannienne, calcul d'Otto. Formellement, étant donnée une énergie E sur $\mathcal{P}_2(\mathbf{R}^d)$, l'équation

$$\partial_t \rho = \operatorname{div}(\rho \nabla E'(\rho))$$

est le flot de gradient (pour W_2) de E . En particulier

$E(\rho) = \int_{\mathbf{R}^d} \rho \log(\rho)$: le flot de la chaleur est le flot de l'entropie pour W_2 .

Jordan-Kinderlehrer et Otto, 1998, l'équation de Fokker-Planck:

$$\partial_t \rho - \Delta \rho - \operatorname{div}(\rho \nabla V) = 0$$

est le flot gradient pour W_2 de

$$E(\rho) = \int_{\mathbf{R}^d} \rho \log(\rho) + \int_{\mathbf{R}^d} V \rho = \int_{\mathbf{R}^d} \rho \log \left(\frac{\rho}{e^{-V}} \right).$$

Résolution rigoureuse via un schéma d'Euler implicite: le schéma JKO. S'adapte à une variété considérable de situations: diffusions non linéaires (Otto, Agueh...), équations d'agrégation, milieux granulaires (Carrillo, Villani, McCann...), équations du quatrième ordre (films minces, Derrida-Lebowitz-Speer-Spohn, McCann, Matthes, Savaré, Toscani...), Chapman-Rubinstein-Schatzman en superconductivité (Ambrosio-Serfaty)....

Beaucoup de résultats: convergence du schéma JKO, caractérisation des flots gradients, unicité, rôle de la convexité par déplacement dans la stabilité, la convergence en temps grand vers l'équilibre etc... Livres de Villani, Ambrosio-Gigli et Savaré, Santambrogio.

Parfois un décalage entre l'avancement de la théorie et les méthodes numériques disponibles.

Plan

- ① Flots gradients: exemples et discrétisation (JKO)
- ② Discrétisation de l'opérateur de Monge-Ampère
- ③ Régularisation entropique

Principe et exemples

Partons d'un flot de gradient dans \mathbf{R}^d (ou un espace de Hilbert):

$$\dot{x}(t) = -\nabla E(x(t)), \quad x(0) = x_0.$$

Schéma d'Euler implicite, étant donné un pas de temps $\tau > 0$,

$$\frac{x^{k+1} - x^k}{\tau} = -\nabla E(x^{k+1}), \quad x^0 = x_0$$

qui est l'équation d'Euler de

$$x^{k+1} \in \operatorname{argmin}_x \left\{ \frac{1}{2\tau} |x - x^k|^2 + E(x) \right\}.$$

Mouvement minimisant à la De Giorgi, même idée mais dans un espace métrique.

Schéma JKO pour l'EDP:

$$\partial_t \rho = \operatorname{div}(\rho \nabla E'(\rho)), \quad \rho|_{t=0} = \rho_0.$$

On part de ρ_0 et on construit par récurrence une suite de mesures de probabilité via

$$\rho^{k+1} \in \operatorname{argmin}_{\rho \in \mathcal{P}_2(\mathbf{R}^d)} \left\{ \frac{1}{2\tau} W_2^2(\rho, \rho^k) + E(\rho) \right\}.$$

On pose $\rho^\tau(t) = \rho^{k+1}$ pour $t \in (k\tau, (k+1)\tau]$. Et on cherche à montrer que quand $\tau \rightarrow 0$, ρ^τ converge vers une solution de l'EDP. La difficulté dépend de E , pour fixer les idées prenons une énergie interne

$$E(\rho) := \int_{\mathbf{R}^d} U(\rho(x)) dx$$

avec U strictement convexe, $U(0) = 0$, $U \geq 0$ (pour simplifier)...

Ce qui correspond à l'équation de diffusion non linéaire:

$$\partial_t \rho = \Delta P(\rho), \quad P(\rho) := \rho U'(\rho) - U(\rho).$$

cas particulier: $U(\rho) = \frac{1}{m-1} \rho^m$, $m > 1$, équation des milieux poreux

$$\partial_t \rho = \Delta \rho^m.$$

Equation d'Euler pour la $k + 1$ -ème étape de JKO:

$$\frac{1}{\tau} (y - S^{k+1}(y)) \rho^{k+1}(y) = -\nabla U'(\rho^{k+1}(y)) \rho^{k+1} = -\nabla P(\rho^{k+1}(y)) \quad (1)$$

avec $S_{\#}^{k+1} \rho^k = \rho^{k+1}$, optimal et donc

$$\int_{k\tau}^{(k+1)\tau} \int_{\mathbf{R}^d} |\nabla P(\rho^\tau)| dx dt = \tau \int_{\mathbf{R}^d} |\nabla P(\rho^{k+1})| \leq W_2(\rho^k, \rho^{k+1})$$

Evidemment, on a

$$W_2^2(\rho^{k+1}, \rho^k) \leq 2\tau \left(E(\rho^k) - E(\rho^{k+1}) \right).$$

De sorte que (avec $N \simeq T/\tau$)

$$\sum_{k=0}^{N-1} W_2^2(\rho^{k+1}, \rho^k) \leq C\tau, \quad (2)$$

D'où une borne $C^{0,1/2}$ en temps

$$W_2(\rho^\tau(t), \rho^\tau(s)) \leq C\sqrt{|t-s| + \tau}. \quad (3)$$

et une estimation $L^1((0, T), \text{BV})$ pour la pression

$$\sum_{k=0}^{N-1} \int_{k\tau}^{(k+1)\tau} \int_{\mathbf{R}^d} |\nabla P(\rho^\tau)| dx dt \leq C(N\tau)^{1/2} \leq C_T. \quad (4)$$

Plus des bornes sur l'énergie et les moments d'ordre 2. Par un argument de type Aubin-Lions-Simon (cf. Rossi-Savaré pour le cas d'un module de continuité en temps donné par une distance plutôt qu'une estimation Sobolev en temps), convergence forte: la famille $\{P(\rho^\tau)\}_\tau$ est relativement compacte dans $L^1((0, T) \times \mathbf{R}^d)$. Convergence forte de ρ^τ et $P(\rho^\tau)$: limites ρ et $P(\rho)$.

Fonction-test $\phi \in C_c^\infty((0, T) \times \mathbf{R}^d)$:

$$\begin{aligned}
 \int_0^T \int_{\mathbf{R}^d} \partial_t \phi \rho^\tau &= - \sum_{k=0}^N \int_{\mathbf{R}^d} \phi(k\tau, \cdot) (\rho^{k+1} - \rho^k) \\
 &= - \sum_{k=0}^N \int_{\mathbf{R}^d} (\phi(k\tau, y) - \phi(k\tau, S^{k+1}(y))) \rho^{k+1}(y) dy \\
 &= - \sum_{k=0}^N \int_{\mathbf{R}^d} \nabla \phi(k\tau, y) \cdot (y - S^{k+1}(y)) \rho^{k+1}(y) dy \\
 &\quad + O(\tau) \text{ (avec (3))}
 \end{aligned}$$

$$\text{(avec (1))} = - \sum_{k=0}^N \tau \int_{\mathbf{R}^d} \Delta \phi(k\tau, x) P(\rho^\tau(x)) dx + O(\tau).$$

Et donc en faisant $\tau \rightarrow 0$, $\partial_t \rho = \Delta P(\rho)$.

S'adapte sous des hypothèses convenables à des énergies plus générales:

$$\begin{aligned}
 E(\rho) := & \int_{\mathbf{R}^d} U(\rho(x))dx + \int_{\mathbf{R}^d} V(x)\rho(x)dx \\
 & + \frac{1}{2} \int_{\mathbf{R}^d \times \mathbf{R}^d} W(x-y)\rho(x)\rho(y)dxdy
 \end{aligned} \tag{5}$$

dont le flot de gradient W_2 s'écrit:

$$\partial_t \rho = \operatorname{div} \left(\rho(\nabla U'(\rho) + \nabla V + \nabla W \star \rho) \right).$$

Stabilité, unicité, contraction pour W_2 , convergence vers l'équilibre: ok si E est convexe par déplacement i.e. $t \mapsto E(\rho_t)$ est convexe le long des géodésiques (V est W convexe et U vérifie la condition de McCann). Terme d'interaction W : équations d'agrégation, milieux granulaires...

Quelques exemples plus compliqués.

Mouvements de foules: Maury, Roudneff-Chupin,
Santambrogio

$$\partial_t \rho = \operatorname{div}(\rho(\nabla V + \nabla p)), \quad p \geq 0, \quad \rho \leq 1, \quad p(1 - \rho) = 0.$$

qui correspond à

$$E(\rho) = \begin{cases} \int_{\mathbf{R}^d} V \rho & \text{si } \rho \leq 1 \\ +\infty & \text{sinon} \end{cases} .$$

Modèles du quatrième ordre: Jüngel, Toscani, Matthes, McCann, Savaré... Equation des films minces

$$\partial_t \rho = -\operatorname{div} \left(\rho \nabla \Delta \rho \right) \text{ correspond à } E(\rho) = \frac{1}{2} \int_{\mathbf{R}^d} |\nabla \rho|^2.$$

Equation de dérive-diffusion quantique de Derrida-Lebowitz-Speer-Spohn:

$$\partial_t \rho = -\operatorname{div} \left(\rho \nabla \left(\frac{\Delta \sqrt{\rho}}{\sqrt{\rho}} + V \right) \right)$$

qui correspond à

$$E(\rho) = \int_{\mathbf{R}^d} |\nabla \sqrt{\rho}|^2 + V \rho.$$

Et maintenant: peut-on utiliser le schéma JKO numériquement en pratique? Résoudre:

$$\inf_{\rho \in \mathcal{P}_2(\mathbf{R}^d)} \left\{ \frac{1}{2\tau} W_2^2(\rho, \rho^k) + E(\rho) \right\}$$

pour un E "raisonnable" (disons de la forme (5)). La difficulté est le terme de transport optimal W_2^2 . Facile en dimension un, car la géométrie est simple:

$$W_2^2(\mu, \nu) = \int_0^1 (F_\mu^{-1}(s) - F_\nu^{-1}(s))^2 ds,$$

moins facile en dimension supérieure.

Discrétisation de l'opérateur de Monge-Ampère

Travail commun avec J.-D. Benamou, Q. Mérigot et E. Oudet.
 Soit X et Y deux ouverts convexes bornés de \mathbf{R}^2 ,
 $\mu \in \mathcal{P}(X) \cap L^\infty(X)$ avec $\mu \geq r > 0$, on souhaite résoudre

$$\inf_{\nu \in \mathcal{P}(Y)} \mathcal{F}(\nu) := W_2^2(\mu, \nu) + E(\nu)$$

avec

$$E(\nu) := \mathcal{U}(\nu) + \mathcal{V}(\nu), \quad \mathcal{U}(\nu) = \begin{cases} \int_Y U(\nu(y)) dy & \text{si } \nu \in L^1, \\ +\infty & \text{sinon} \end{cases}$$

avec

$$\mathcal{V}(\nu) := \int_Y V(y) d\nu(y) + \int_{Y \times Y} W(x - y) \nu(dx) \otimes \nu(dy).$$

Grâce au théorème de Brenier, le problème est équivalent à

$$\inf_{u \text{ convexe}, \nabla u \in \bar{Y}} \mathcal{J}(u) := \int_X |x - \nabla u(x)|^2 \mu(x) dx + E(\nabla u_{\#} \mu) \quad (6)$$

on a

$$\begin{aligned} \mathcal{V}(\nabla u_{\#} \mu) &= \int_X V(\nabla u(x)) \mu(x) dx \\ &+ \int_{X \times X} W(\nabla u(x) - \nabla u(x')) \mu(dx) \mu(dx'). \end{aligned}$$

Pour $\mathcal{U}(\nabla u_{\#}\mu)$ c'est plus compliqué: si u est régulière, $\nu = \nabla u_{\#}\mu$ est liée à u et μ par l'équation de Monge-Ampère:

$$\mu(x) = \det(D^2u(x))\nu(\nabla u(x))$$

ainsi

$$\begin{aligned}\mathcal{U}(\nabla u_{\#}\mu) &= \int_X U(\nu(\nabla u(x))) \det D^2u(x) dx \\ &= \int_X U\left(\frac{\mu(x)}{\det D^2u(x)}\right) \det D^2u(x) dx.\end{aligned}$$

Sous la condition de Mc Cann:

$$r > 0 \mapsto r^d U(r^{-d}) \text{ convexe décroissante}$$

avec la concavité de $A \in S_d^+ \mapsto \det(A)^{1/d}$,

u convexe $\mapsto \mathcal{U}(\nabla u_{\#}\mu)$ est convexe.

Si u est convexe mais pas nécessairement régulière, mesure de Monge-Ampère:

$$\text{MA}[u](B) := |\partial u(B)|.$$

On note $\mathcal{K} := \{u : X \rightarrow \mathbf{R} \text{ convexe}\}$ et $\mathcal{K}_Y := \{u \in \mathcal{K}, \nabla u \in Y \text{ p.p.}\}$, on suppose que V et W sont convexes, que $U(0) = 0$ et U vérifie la condition de McCann, on souhaite discrétiser le problème convexe

$$\inf_{u \in \mathcal{K}_Y} \mathcal{J}(u)$$

On se donne $P \subset X$ fini et une discrétisation $\mu_P := \sum_{p \in P} \mu_p \delta_p$ de μ . Polyèdres: $\mathcal{K}_Y(P) := \{u|_P, u \in \mathcal{K}_Y\}$, et

$$\mathcal{K}_Y^G(P) := \{(u, G_u) \in (\mathbf{R} \times Y)^P : u(q) - u(p) \geq G_u(p) \cdot (q - p), \forall (p, q)\}$$

(cf. Ekeland et Moreno-Bromberg) $|P|^2$ contraintes linéaires.

Si $(u, G_u) \in \mathcal{K}_Y^G(P)$, on a

$$W_2^2(\mu_P, G_u \# \mu_P) = \sum_{p \in P} |p - G_u(p)|^2 \mu_p,$$

et

$$\mathcal{V}(G_u \# \mu_P) = \sum_{p \in P} V(G_u(p)) \mu_p + \sum_{(p, q) \in P^2} W(G_u(p) - G_u(q)) \mu_p \mu_q$$

en revanche on ne peut utiliser la mesure image discrète $G_u \# \mu_P$ pour le terme d'énergie interne.

Discrétisation de la mesure de Monge-Ampère et mesure image absolument continue. Soit $u: P \rightarrow \mathbf{R}$ notons

$$u_{\mathcal{K}_Y} := \sup\{v \in \mathcal{K}_Y : v \leq u \text{ sur } P\}$$

et

$$\text{MA}_Y[u](p) := |V_p|, \quad V_p := \partial u_{\mathcal{K}_Y}(p) := Y \cap \text{Lag}_P^u(p)$$

où la cellule de Laguerre $\text{Lag}_P^u(p)$ est donnée par

$$\text{Lag}_P^u(p) := \{y \in \mathbf{R}^d : u(q) - u(p) \geq y \cdot (q - p), \forall q \in P\}$$

On a:

- $u \in \mathcal{K}_Y(P)$ ssi pour tout $p \in P$, $V_p \neq \emptyset$ auquel cas $\{V_p\}_p$ est une partition de Y en cellules convexes,
- u est dans l'intérieur de $\mathcal{K}_Y(P)$ ssi $\text{MA}_Y[u](p) > 0$ pour tout p .

Soit $u \in \mathcal{K}_Y(P)$ intérieure, on définit $G_u^{\text{ac}} \# \mu_p$ par

$$G_u^{\text{ac}} \# \mu_p := \sum_{p \in P} \frac{\mu_p}{\text{MA}_Y[u](p)} \mathbf{1}_{\partial u_{\mathcal{K}_Y}(p)}.$$

Lemme 1 *Si U vérifie la condition de McCann alors*

$$u \mapsto \mathcal{U}(G_u^{\text{ac}} \# \mu_P) = \sum_{p \in P} U\left(\frac{\mu_p}{\text{MA}_Y[u](p)}\right) \text{MA}_Y[u](p)$$

est convexe sur l'intérieur de $\mathcal{K}_Y(P)$.

Si en outre U est surlinéaire $\lim_{r \rightarrow \infty} U(r)/r = +\infty$, \mathcal{U} est une barrière pour la contrainte de convexité:

$$\mathcal{U}(G_u^{\text{ac}} \# \mu_P) < +\infty \Rightarrow u \in \text{int}(\mathcal{K}_Y(P)).$$

On a remplacé $|P|^2$ contraintes linéaires en $|P|$ contraintes convexes!

Problème discret:

$$\inf_{(u, G_u) \in \mathcal{K}_Y^G(P)} W_2^2(\mu_P, G_u \# \mu_P) + \mathcal{U}(G_u^{\text{ac}} \# \mu_P) + \mathcal{V}(G_u \# \mu_P)$$

- Si V et W convexes, U vérifie la condition de McCann, c'est un problème convexe,
- il semble nécessaire d'utiliser deux notions de sous-gradients, l'une discrète, l'autre continue (car $\mathcal{V}(G_u^{\text{ac}} \# \mu_P)$ n'est pas convexe).

Convergence, si on suppose que:

- \mathcal{V} est continue et \mathcal{U} est sci (pour W_2),
- U est convexe et surlinéaire à l'infini, $U(0) = 0$,
- μ_{P_n} converge faiblement vers μ ,

alors

Théorème 1 *La suite $G_{u_n \# \mu_{P_n}}^{\text{ac}}$ associée à des solutions du problème discrétisé est une suite minimisante pour \mathcal{F} .*

En pratique, il faut reconstruire le diagramme de Laguerre et sa triangulation duale: problème standard de géométrie algorithmique, calcul en temps $|P| \log(|P|)$. Jacobiennes et hessiennes de l'opérateur de Monge-Ampère discrets sont très creuses: on peut utiliser une méthode de Newton.

Avantages:

- énergie interne agit comme une barrière à la convexité au niveau discret, rapide et efficace (Newton),
- reconstruction d'une mesure absolument continue et convergence forte, coût quasiment linéaire en le nombre de points,
- peut s'adapter à la minimisation de certaines fonctionnelles en $\det(D^2u)$ (Problème de Plateau affine, Abreu...).

Inconvénients:

- assez complexe à mettre en oeuvre (compétences en géométrie algorithmique),
- difficile à adapter en pratique aux fonctionnelles non géod. convexes.

Régularisation entropique

Résoudre un problème de transport optimal est généralement très coûteux, même pour des mesures empiriques, le problème approché en rajoutant ε fois une entropie:

$$\inf_{\gamma \in \Pi(\mu, \nu)} \langle c, \gamma \rangle + \varepsilon \int_{\mathbf{R}^d \times \mathbf{R}^d} \gamma \log \gamma$$

avec $\varepsilon > 0$ et μ et ν d'entropie finie est beaucoup plus simple. C'est un problème de projection pour la divergence de Kullback-Leibler

$$\inf_{\gamma \in \Pi(\mu, \nu)} \text{KL}(\gamma | \gamma^\varepsilon) := \int_{\mathbf{R}^d \times \mathbf{R}^d} \gamma \log \left(\frac{\gamma}{\gamma^\varepsilon} \right), \quad \gamma^\varepsilon = e^{-\frac{c}{\varepsilon}}.$$

Problème discret

$$\inf_{\gamma \in C_1 \cap C_2} \sum_{ij} \gamma_{ij} \log \left(\frac{\gamma_{ij}}{\theta_{ij}} \right) \quad (7)$$

avec

$$C_1 := \left\{ \gamma : \sum_j \gamma_{ij} = \mu_i, \forall i \right\}, \quad C_2 := \left\{ \gamma : \sum_i \gamma_{ij} = \nu_j, \forall j \right\}$$

Les projections de θ pour KL sur C_1 et C_2 sont totalement explicites:

$$\text{proj}_{C_1}^{\text{KL}}(\theta)_{ij} = \frac{\theta_{ij} \mu_i}{\sum_l \theta_{il}}, \quad \text{proj}_{C_2}^{\text{KL}}(\theta)_{ij} = \frac{\theta_{ij} \nu_j}{\sum_l \theta_{lj}},$$

Projection alternées (aka IPFP, Iterative Proportional Fitting Procedure):

$$C_{2n+1} = C_1, C_{2n} = C_2, \gamma^n = \text{proj}_{C_n}^{\text{KL}}(\gamma^{n-1}).$$

Alors γ^n converge vers la solution de (7) (contraction pour la métrique de Hilbert). Calculs hautement parallélisables, seulement $2I$ coefficients à stocker à chaque étape.

Idée ancienne (remonte à Schrödinger), énorme littérature:

- En stats et probas (Csisczar, Dykstra, Léonard, Föllmer, Rüschendorf...),
- En optimisation (Bregman, Bauschke, Combettes, Lewis, Cominetti, San-Martin...).

Plus récente pour le transport optimal et ses applications: Galichon, Salanié (économétrie) et Cuturi (machine learning, complexité lien avec Sinkhorn). S'adapte facilement aux problèmes multimarges, au transport partiel, aux barycentres au sens de W_2 etc... cf. Benamou, C., Cuturi, Nenna et Peyré.

Première application. Barycentres pour W_2 , on se donne $(\mu_1, \dots, \mu_k) \in \mathcal{P}_2(\mathbf{R}^d)^k$ des poids $\lambda_1, \dots, \lambda_k$ et on cherche à interpoler via

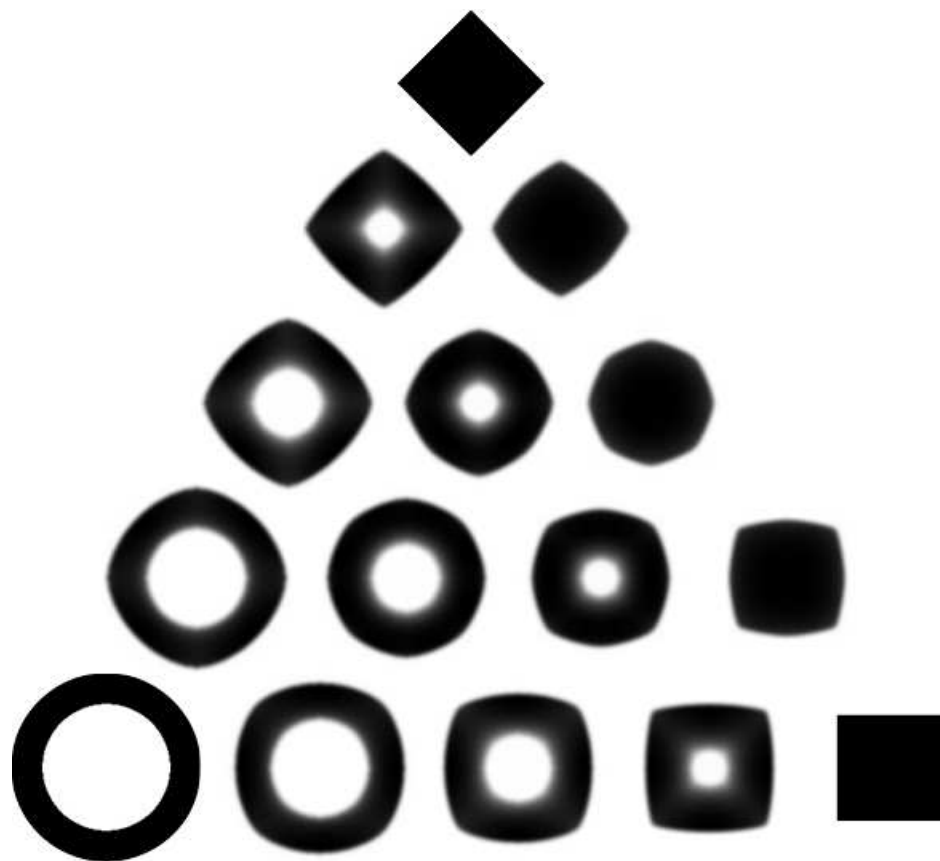
$$\inf_{\mu \in \mathcal{P}_2(\mathbf{R}^d)} \sum_{i=1}^k \lambda_i W_2^2(\mu_i, \mu)$$

une solution est un barycentre au sens de W_2 (moyenne de Fréchet). Existence, unicité, régularité, cas des gaussiennes etc... cf C., Agueh. Le problème est équivalent à

$$\inf_{(\gamma_1, \dots, \gamma_k) \in \mathcal{P}_2(\mathbf{R}^{2d})} \sum_{i=1}^k \int_{\mathbf{R}^{2d}} |x_i - x|^2 d\gamma_i(x_i, x)$$

sous les contraintes que la première marge de γ_i est μ_i et la seconde ne dépend pas de i (c'est le barycentre).

Régularisation entropique se résout rapidement par IPFP:



cf. Benamou, C., Cuturi, Nenna et Peyré.

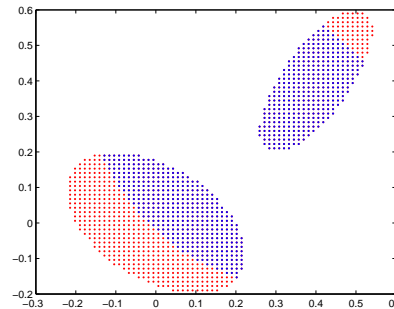
Application 2: transport partiel. Transport partiel, deux mesures positives, μ et ν (pas de même masse a priori), $m \leq \min(\mu(X), \nu(Y))$ et l'on s'intéresse à

$$\inf_{\gamma \in \Pi_m^-(\mu, \nu)} \langle c, \gamma \rangle$$

avec

$$\Pi_m^-(\mu, \nu) := \{ \gamma \in \mathcal{P}(X \times Y) : \pi_{1\#}\gamma \leq \mu, \pi_{2\#}\gamma \leq \nu, \gamma(X \times Y) = m \}$$

intersection de trois convexes sur lesquels les projections KL sont explicites.



G. Peyré: idée d'utiliser cette régularisation entropique pour JKO:

$$\rho^{k+1} \in \operatorname{argmin}_{\rho \in \mathcal{P}_2(\mathbf{R}^d)} \left\{ \frac{1}{2} W_2^2(\rho, \rho^k) + \tau E(\rho) \right\} \quad (8)$$

approché par

$$\rho^{k+1} \in \operatorname{argmin}_{\rho \in \mathcal{P}_2(\mathbf{R}^d)} \left\{ \frac{1}{2} K_\varepsilon(\rho, \rho^k) + \tau E(\rho) \right\} \quad (9)$$

où $\varepsilon > 0$ et

$$K_\varepsilon(\mu, \nu) := \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbf{R}^d \times \mathbf{R}^d} |x - y|^2 \gamma(x, y) dx dy \\ + \varepsilon \int_{\mathbf{R}^d \times \mathbf{R}^d} \gamma(x, y) \log(\gamma(x, y)) dx dy.$$

Deux petits paramètres ε et τ , comment les relier pour avoir convergence? Prenons:

$$E(\rho) := \int_{\mathbf{R}^d} U(\rho(x))dx + \int_{\mathbf{R}^d} v(x)\rho(x)dx$$

avec (par exemple) $U(\rho) = \rho \log(\rho)$ ou ρ^m , $m > 1$, $v \geq 0$ et régulière et $\rho_0 \in \mathcal{P}_2(\mathbf{R}^d)$, $E(\rho_0) < +\infty$. Le flot de gradient de E correspond donc à

$$\partial_t \rho = \Delta P(\rho) + \operatorname{div}(\rho \nabla v), \quad \rho|_{t=0} = \rho_0, \quad (10)$$

avec

$$P(\rho) = \rho U'(\rho) - U(\rho).$$

Equation d'Euler-Lagrange de (9): semblable à la version non régularisée un terme en $\frac{\varepsilon}{\tau} \Delta \rho^{k+1}$ en plus et surtout γ^{k+1} optimal pour $K_\varepsilon(\rho^k, \rho^{k+1})$ et non plus $W_2^2(\rho^k, \rho^{k+1})$.

Convergence:

Théorème 2 (*C., Duval, Peyré, Schmitzer*) *Le schéma (9) converge vers une solution de (10) quand $\varepsilon, \tau \rightarrow 0$ dès que $\varepsilon |\log(\varepsilon)| = O(\tau^2)$.*

Avantages:

- très rapide, facile à implémenter et souple (on peut prendre un coût non quadratique...), méthode "générique",
- facile à utiliser, très adapté aux applications en machine learning.

Inconvénients:

- en pratique pour les flots gradients $\varepsilon \log \varepsilon = O(\tau^2)$: très peu de pas de temps, on ne peut pas prendre ε vraiment petit.

Un peu de publicité:

- Application de l'IPFP à la dynamique des fluides (cf. exposé de Luca Nenna)
- Une autre méthode eulérienne: Benamou-Brenier et Lagrangien augmenté, s'adapte aux systèmes (cf. exposé de Maxime Laborde).