Introduction au domaine de recherche Transport optimal et calcul des variations

Jules Candau-Tilh
Sous la direction de Guillaume Carlier

Octobre 2020

Résumé

Le problème du transport optimal a été formalisé pour la première fois par le mathématicien français Gaspard Monge en 1781. La théorie connaît un regain d'intérêt après le développement de la formulation duale du transport optimal par le mathématicien et économiste russe Leonid Kantorovich en 1942. Depuis la fin des années 1990 et les travaux de Yann Brenier, le transport optimal a trouvé des applications dans des domaines aussi variés que l'équation de Boltzmann (Villani, Otto), les inégalités isopérimétriques (Figalli), la géométrie riemannienne (McCann) ou encore l'analyse d'images (Peyré, Bonneel). Plus précisément, nous verrons que cette théorie permet de mesurer des distances entre des mesures de probabilités, ce qui fait apparaître le problème de la géodésie des espaces de probabilité.

Table des matières

1 Historique du transport optimal

Cette présentation s'inspire du cours donné par Luigi Ambrosio à la Scuola Normale Superiore [3], auquel l'auteur a assisté en 2019.

1.1 Monge et le problème des déblais et des remblais

Dans son célèbre mémoire de 1781 "Sur la théorie des déblais et des remblais" [11], le mathématicien Gaspard Monge pose pour la première fois le problème du transport optimal. La question est celle du déplacement d'un volume de terre localisé dans un "déblai" vers un espace d'arrivée appelé "remblai", déplacement qui doit se faire de manière à minimiser la quantité de terre transportée. Dans ce mémoire, la solution au problème ainsi formulé n'est pas donnée, malgré les affirmations optimistes de Monge. Il faudra en effet attendre plus de deux siècles pour qu'en 2003 le mathématicien italien Luigi Ambrosio corrige la preuve donnée par N. V. Sudakov en 1979 [14] et formule une réponse définitive à la version moderne du problème de Monge [1].

Mathématiquement parlant, on se donne deux espaces X et Y polonais (i.e. munis d'une distance les rendant complets et séparables), ainsi que deux mesures de probabilité $\mu \in \mathcal{P}(X)$ et $\nu \in \mathcal{P}(Y)$ représentant respectivement les positions de départ et d'arrivée de la masse transportée. Soit alors un coût de transport borélien $c : X \times Y \longrightarrow [0, +\infty]$, il s'agit de résoudre le problème suivant :

$$\inf_T \left\{ \int_X c(x, T(x)) \mathrm{d}\mu, \; T : X \to Y \text{ borélienne}, T_\# \mu = \nu \right\}. \tag{M}$$

La condition $T_\# \mu = \nu$ signifie que la fonction T doit "pousser en avant" la mesure μ "sur" la mesure ν : le déblai s'envoie sur le remblai. Rigoureusement, pour tout borélien B, on a

$$\nu(B) = T_\# \mu(B) = \mu(T^{-1}(B)).$$

Cette formulation du transport optimal, dite de Monge, est plutôt intuitive, notamment en considérant des mesures discrètes.

Exemple 1.1. (Le cas discret.)

On considère les espaces $X = (x_1, x_2, \ldots, x_n)$ et $Y = (y_1, y_2, \ldots, y_n)$ munis de la distance uniforme, et où les $(x_i)_i$ et les $(y_i)_i$ sont distincts deux à deux. On construit alors les mesures de probabilités

$$\mu = \sum_{i=1}^N \frac{1}{N} \delta_{x_i} \text{ et } \quad \nu = \sum_{i=1}^N \frac{1}{N} \delta_{y_i}.$$

Pour toute fonction $T : X \to Y$, on a

$$T_\# \mu = \frac{1}{N} \sum_{i=1}^N \delta_{T(x_i)},$$

de sorte que T est admissible pour le transport optimal, i.e. $T_{\#}\mu = \nu$ si et seulement si elle est bijective. Ainsi, pour déterminer le transport optimal, il faut et il suffit de trouver la meilleure permutation T parmi les $n!$ choix possibles. Numériquement cependant, évaluer le coût de chaque permutation peut rapidement s'avérer sans issue.

Exemple 1.2. (Le réarrangement monotone.) L'autre exemple fondamental de la formulation de Monge est celui du cas continu en dimension un. Pour simplifier, on considère deux probabilités $\mu, \nu << \mathcal{L}^1$ telles que leur support soit $\mathbb{R}$ tout entier. On note alors F_μ et F_ν les fonctions de répartition de μ et ν, qui sont continues, et la condition $T_{\#}\mu = \nu$ s'écrit :

$$\forall x \in \mathbb{R},\ F_\nu(T(x)) = F_\mu(x).$$

Comme $\mathrm{supp}(\nu) = \mathbb{R}$, F_ν est continue et strictement croissante, elle est donc inversible, et nécessairement

$$T = F_\nu^{-1}(F_\mu).$$

On a alors construit une fonction de transport continue et strictement monotone, c'est pourquoi on parle de réarrangement monotone. L'unicité est ici évidente par construction. Dans le cas ou $\mathrm{supp}(\nu) \neq \mathbb{R}$, on doit travailler avec le pseudo-inverse de F_ν, mais l'idée de la preuve reste sensiblement la même.

Le concept du réarrangement monotone s'avère être fondamental dans la théorie du transport optimal, car il donne une preuve constructive de la meilleure fonction de transport. Beaucoup de constructions en dimension supérieure reposent en effet sur un réarrangement monotone dans chaque direction : on pense notamment à la fonction de Knothe-Rosenblatt [5].

1.2 La formulation duale de Kantorovich

Un des principaux problèmes de la formulation dite de Monge du transport optimal est qu'une solution n'existe pas toujours, même dans des cas apparemment simples. Prenons par exemple un transport dans $\mathbb{R}$, de δ_0 vers $(\delta_{-1} + \delta_1)/2$: physiquement, cela correspond juste à diviser une masse en deux parts égales et à les amener en deux points différents. Mais, dans la formulation de Monge, cela signifierait que la carte de transport T prend deux valeurs différentes à l'arrivée, ce qui est impossible pour une fonction.

C'est là que l'approche [9] du mathématicien et économiste russe Léonid Kantorovich prend tout son sens : il introduit le concept de plan de transport, ce qui revient à chercher une mesure dans $\mathcal{M}(X \times Y)$ dont les marginales sont les mesures de départ et d'arrivée. Concrètement, on se donne $\mu, \nu \in \mathcal{P}(X) \times \mathcal{P}(Y)$ et on définit l'ensemble des plans de transport entre μ et ν comme :

$$\Gamma(\mu, \nu) = \big\{\pi \in \mathcal{M}(X \times Y),\ (p_X)_{\#}\pi = \mu,\ (p_Y)_{\#}\pi = \nu\big\}.$$

Par suite, si on se donne une fonction de coût borélienne $c : X \times Y \to [-\infty, +\infty]$, le problème du transport optimal s'écrit :

$$\inf\left\{ \int_{X\times Y} c(x,y)\,\mathrm{d}\pi,\ \ \pi\in\Gamma(\mu,\nu)\right\}. \tag{K}$$

Cette formulation présente beaucoup d'avantages par rapport à celle de Monge. Le concept de plan de transport est en effet plus souple que celui de fonction de transport. En effet, si π et π' sont deux plans de transport admissibles, alors $\pi'' = (\pi+\pi')/2$ en est un également. Cela est d'une grande aide dans les preuves d'unicité des principaux théorèmes. Par ailleurs, les plans de transport généralisent les cartes de transport : si T est une carte de transport, alors il est immédiat que $\pi = (\mathrm{Id}\times T)_{\#}\mu$ est un plan de transport. Un des autres avantages est que l'existence et l'unicité d'un plan de transport optimal peut s'obtenir pour des hypothèses très faibles, comme le montre le théorème suivant.

Théorème 1.3. (Existence et unicité du plan de transport optimal).

Supposons que X et Y sont complets et séparables et que le coût c est positif, fini partout et continu. Si on se donne $\mu\in\mathcal{P}(X)$, $\nu\in\mathcal{P}(Y)$, alors il existe un unique plan de transport optimal entre μ et ν.

Preuve. (Idée de preuve de l'existence et l'unicité).

Partant de ces hypothèses, il est facile d'obtenir que

$$\mathcal{C} : \pi\in\Gamma(\mu,\nu)\mapsto \int_{X\times Y} c\,\mathrm{d}\pi,$$

est semi-continue inférieurement pour la topologie faible sur $\mathcal{M}(X\times Y)$, c'est-à-dire en dualité avec $\mathcal{C}_b(X\times Y)$. L'idée de la preuve repose sur l'approximation monotone de c par des fonctions lipschitziennes bornées, pour lesquelles l'affirmation se vérifie facilement. Par ailleurs, en utilisant le lemme d'Ulam, on montre facilement que $\Gamma(\mu,\nu)\subset\mathcal{P}(X,Y)$ est un compact, et donc que l'infimum dans le problème de Kantorovich est un minimum.

Pour construire des solutions au problème de Kantorovich et en établir l'unicité, il y a plus à faire. Le tout repose sur des concepts d'analyse convexe, et en particulier sur une généralisation de la convexité : la c-convexité. Étant toujours donné notre coût c, on dit qu'une fonction $\phi : X\to\mathbb{R}\cup\{-\infty\}$ est c-concave s'il existe $\psi : Y\to\mathbb{R}\cup\{-\infty\}$ telle que

$$\forall x\in X,\ \phi(x) = \inf_{y\in Y}\{c(x,y)-\psi(y)\}$$

On remarque que pour $X=Y=\mathbb{R}^d$ et $c(x,y)=|x-y|^2$, on retrouve bien le concept de la conjuguée convexe d'une fonction, aussi appelée transformée de Legendre. Le dernier concept intervenant dans la construction de l'unicité des solutions est celui de la c-cyclicité monotone : on qu'une ensemble $\Gamma\subset X\times Y$ est c-cycliquement monotone si pour tout $N\geq 1$ et $(x_i,y_i)_{1\leq i\leq n}\in\Gamma^N$,

$$\sum_{i=1}^{N} c(x_i,y_i)\leq\sum_{i=1}^{N} c(x_i,y_{\sigma(i)})\ \forall\sigma\in\mathfrak{S}_N$$

Formellement, c'est le genre d'ensemble parfait pour le transport monotone, puisque tout plan de transport dont le support serait dans un Γ c-cycliquement monotone devrait être optimal. En effet, toute permutation ne ferait qu'augmenter le coût de transport.

Le lien entre l'analyse convexe et le problème de Kantorovich se fait alors via le théorème de dualité de Kantorovich-Rubinstein, que l'on admettra :

$$\inf_{(K)} = \sup_{\phi,\psi\in I_c} \left\{ \int_X \phi \, \mathrm{d}\mu + \int_Y \psi \, \mathrm{d}\nu \right\},$$

où $I_c = \{(\phi,\psi) \in \mathrm{Lip}_b(X) \times \mathrm{Lip}_b(Y),\ \phi(x) + \psi(y) \leq c(x,y)\}$.

L'idée de la preuve est alors comme suit. On sait qu'il existe un plan de transport π optimal. Un théorème qu'on admet nous donne que son support est c-cycliquement monotone. On peut alors construire une fonction ϕ c-concave telle que $\mathrm{supp}(\pi) \subset \mathrm{graph}(\partial^c\phi)$ où $\partial^c\phi$ est l'ensemble des c-sous différentielles de ϕ. C'est le théorème de Rockafellar généralisé. La définition exacte est $y \in {}^c\partial\phi(x)$ si et seulement si

$$c(x,y) - \phi(x) \leq c(x',y) - \phi(x') \, \forall x' \ \in X.$$

Donc en réarrangant les termes, il faut que

$$\forall x' \in X,\ \phi(x') \leq c(x',y) - c(x,y) + \phi(x)\ \forall (x,y) \in \mathrm{supp}(\pi).$$

Mais en réitérant cette inégalité N fois, on voit qu'il faut que

$$\phi(x) \leq c(x,y_N) - c(x_N,y_N) + c(x_N,y_{N-1}) - c(x_{N-1},y_{N-1}) + \cdots + c(x_1,y_0) - c(x_0,y_0) + \phi(0) \tag{ineq}$$

On définit alors ϕ comme étant l'infimum sur N du membre de droite de l'inégalité, en posant $\phi(0) = 0$. Le but du jeu devient donc de montrer qu'une telle fonction existe, nécessairement elle vérifiera $\phi(x) + \phi(y) = c(x,y)$ sur tout $\mathrm{supp}(\pi)$. On dit que π est concentré sur un ensemble contact pour c. On a alors

$$\inf_{(K)} = \int_{X\times Y} c \, \mathrm{d}\pi = \int_{X\times Y} (\phi(x) + {}^c\phi(y)) \, \mathrm{d}\pi = \int_X \phi(x) \, \mathrm{d}\mu + \int_Y {}^c\phi \, \mathrm{d}\nu,$$

de sorte qu'on a résolu le problème, et on peut prendre $\pi = (\mathrm{Id} \times \phi)_{\#}\mu$ comme plan de transport optimal.

□

1.3 La fonction de transport de Brenier

Dans les années 1990, en cherchant des factorisations polaires de champs de vecteurs définis sur des compacts [4], le mathématicien français Yann Brenier parvient à exhiber une fonction qui caractérise entièrement le transport optimal de $\mathbb{R}^n$ dans $\mathbb{R}^n$ pour le coût quadratique $c(x,y) = |x-y|^2/2$.

Théorème 1.4. (Brenier, Knott-Smith). *Soit μ, ν deux probabilités sur $\mathbb{R}^n$ de moment quadratique fini telles que $\mu \ll \mathcal{L}^n(\mathbb{R}^n)$, et tel que $c(x,y) = |x-y|^2/2$ soit la distance quadratique sur $\mathbb{R}^n$. Alors*

1. *$\exists! \, \pi \in \Gamma(\mu,\nu)$ optimal, donné par $\pi = (Id \times \nabla\phi)_{\#}\mu$, où ϕ est une fonction convexe semi-continue inférieurement dérivable μ-presque partout.*
2. *Réciproquement, si on se donne $T = \nabla\phi$ où ϕ est convexe semi-continue inférieurement avec $|\nabla\phi| \in L^2(\mathbb{R}^n)$, alors T est une fonction de transport optimale de μ vers $\nu = T_{\#}\mu$.*
3. *Si $\nu \ll \mathcal{L}(\mathbb{R}^n)$, et en notant $T^{\mu\to\nu}, T^{\nu\to\mu}$ les cartes de transport optimal de μ vers ν et de ν vers μ respectivement, on a que :*

$$\begin{aligned} T^{\mu\to\nu} \circ T^{\nu\to\mu} &= \ Id \ \nu\text{-p.p.}, \\ T^{\nu\to\mu} \circ T^{\mu\to\nu} &= \ Id \ \mu\text{-p.p.}. \end{aligned}$$

Preuve. (Idée de preuve du théorème de Brenier-Knott-Smith).

Pour l'existence, on utilise le théorème précédent, qui nous donne un plan de transport π concentré sur un ensemble c-cycliquement monotone Γ, lui-même inclus dans graph$(\partial^c\phi)$ où ϕ est c-concave. Puis sur $\Omega = p_X(\text{graph}({}^c\partial\phi))$, on définit la fonction convexe

$$\psi : x \mapsto \frac{1}{2}|x|^2 - \phi(x).$$

Le point crucial est alors d'invoquer le théorème d'Alexandrov, qui nous donne que toute fonction convexe est presque partout deux fois dérivable, i.e. un développement de Taylor a l'ordre deux avec une erreur quadratiquement contrôlée. On écrit alors :

$$\nabla_x\big(c(x,y) - \phi(x)\big)(x) = (x-y) - \nabla_x\phi(x) = 0$$

Donc $(x,y) \in \Gamma$ donne que

$$y = x - \nabla\phi(x) = \nabla\big(\frac{1}{2}|x|^2 - \phi(x)\big) = \nabla\psi(x),$$

de sorte que l'on a bien exhibé notre carte de transport $T = \nabla\phi$. Concernant l'unicité, si on a deux plans de transport optimaux π et π', alors leur moyenne $\pi'' = (\pi + \pi')/2$ est aussi optimale. Or, chaque plan optimal est associé à une fonction de transport T, on peut alors montrer que nécessairement celle de π'' doit vérifier $T''(x) = T(x) = T'(x)$ presque partout.

Les preuves des autres points n'apportent plus d'idées nouvelles, il s'agit d'utiliser intelligemment les théorèmes d'unicité fournis par l'analyse convexe de ce qui précède. □

En 2001, le mathématicien canadien Robert McCann va encore plus loin et généralise le théorème de Brenier dans le contexte des variétés riemanniennes [10]. On considère donc une variété riemannienne lisse M, compacte sans bord et de dimension n. On la munit de la distance

$$d_M(x,y) = \min\left\{ \int_{t=0}^{1} |\dot\gamma(t)|^2 \,\mathrm{d}t,\ \gamma(0) = x,\ \gamma(1) = y,\ \gamma \in \mathcal{C}^1([0,1], M) \right\},$$

et de la forme volume $\text{vol}_M \in \mathcal{M}_+(M)$ définie comme la mesure de Hausdorff n-dimensionnelle induite par d_M. Il est bien connu que (M, d_M) est un espace géodésique, i.e. pour tout couple de points il existe au moins une courbe minimisant le trajet parcouru entre les deux points. Alors tout géodésique γ entre x et y peut se représenter par

$$\gamma(t) = \exp_x(t\xi), \text{ où } \xi = \gamma'(0) \text{ vérifie } |\xi| = d_M(x,y),$$

et où $\exp_x$ désigne l'application exponentielle de T_xM dans M. La beauté du théorème de McCann réside dans l'établissement d'un parallèle entre l'espace géodésique (M, d_M) et l'espace $(\mathcal{P}(M))$, qu'on peut en fait munir d'une distance géodésique.

Théorème 1.5. (McCann). *Soit $\mu, \nu \in \mathcal{P}(M)$ avec $\mu \ll \mathrm{vol}_M$ et $c = d_M^2/2$. Alors il existe un unique plan de transport optimal π entre μ et ν, qui est induit par une fonction T. De plus, il existe une fonction c-concave et Lipschitz $\phi : M \to \mathbb{R}$ telle que*

$$T(x) = \exp_x(-\nabla\phi(x)) \text{ pour } \mu\text{-presque tout } x.$$

2 Applications et résultats-clés du transport optimal

2.1 Les inégalités isopérimétriques

Le problème isopérimétrique, à savoir celui de déterminer la surface minimale que peut avoir un ensemble de volume donné, remonte à plusieurs millénaires. Cependant, aucune solution rigoureuse ne fut obtenue avant le XIXème siècle et les travaux de Minkowski. Aujourd'hui, à la suite des avancées de Lévy et Gromov, l'obtention de nouvelles inégalités isopérimétriques fait l'objet d'une recherche intense. En 2010, les mathématiciens italiens Figalli, Maggi et Pratelli combinent la preuve de Gromov de l'inégalité isopérimétrique et le théorème de Brenier sur le transport optimal pour obtenir un résultat quantitatif et très général sur l'inégalité isopérimétrique anisotrope [6]. On donne ici une version très simplifiée du théorème.

Théorème 2.1. *Soit B la boule unité de $\mathbb{R}^n$ et E un ouvert à bord $\mathcal{C}^1$ de $\mathbb{R}^n$ tel que $\mathcal{L}^n(E) = \mathcal{L}^n(B)$. Alors $\mathcal{H}^{n-1}(\partial B) \leq \mathcal{H}^{n-1}(\partial E)$, où $\mathcal{H}^{n-1}$ désigne la mesure de Hausdorff de dimension $(n-1)$, c'est-à-dire la mesure surfacique naturelle dans $\mathbb{R}^n$.*

Preuve. (Idée de preuve de l'inégalité isopérimétrique).

On définit

$$\mu := \frac{1}{\omega_n}\mathcal{L}^n \llcorner E, \quad \nu := \frac{1}{\omega_n}\mathcal{L}^n \llcorner B,$$

les mesures de Lebesgue unitaires restreintes à E et B respectivement. Par le théorème de Brenier, on dispose alors d'une fonction de transport $T = \nabla\phi$, où ϕ est convexe. Pour simplifier les idées, on supposera que T est régulière jusqu'au bord de E. Par le théorème de Cauchy-Schwarz $\nabla T = \nabla^2\phi$ est une matrice symétrique réelle donc diagonalisable. De plus, la condition $T_\#\mu = \nu$ entraîne facilement que $\det \nabla T(x) = 1$ pour tout $x \in E$. Par l'inégalité arithmético-géométrique appliquées aux valeurs propres de ∇T, on a :

$$1 = (\det(\nabla T))^{1/n} = (\lambda_1 \, \ldots \, \lambda_n)^{1/n} \leq \frac{1}{n}(\lambda_1 + \cdots + \lambda_n) = \frac{1}{n}\mathrm{div}(T).$$

On applique alors le théorème de divergence de Green-Ostrogradski :

$$\mathscr{H}^{n-1}(\partial B) = n\mathcal{L}^n(E) \leq \int_E \operatorname{div}(T)\,\mathrm{d}x = \int_{\partial E} T \cdot \nu \,\mathrm{d}\mathscr{H}^{n-1} \leq \mathscr{H}^{n-1}(\partial E),$$

la dernière inégalité découlant du fait que T prend ses valeurs dans la boule unité.

□

2.2 Transport optimal et courbure de Ricci

Pour cette partie, on pourra se référer à l'excellente présentation de Figalli et Villani donnée dans [7]. De manière générale, le terme de courbure s'emploie pour parler d'un invariant qui quantifie à quel point un espace diffère localement de l'espace euclidien, c'est-à-dire plat. Le cadre le plus naturel pour étudier la courbure est celui des variétés riemanniennes, qui sont les exemples les plus simples d'espaces non plats mais ressemblant quand même localement à des espaces euclidiens. Si on se donne M une variété riemannienne complète et connectée équipée d'une métrique g, la courbure de Ricci, noté Ric, est un objet qui à chaque point $x \in M$ associe une forme quadratique Ric_x définie sur l'espace tangent T_xM. L'étude de la courbure de Ricci des variétés riemanniennes peut fournir énormément d'informations sur leur structure. On se donne par exemple une courbe $\gamma : t \in [0,1] \to M$ et un ξ un champ de vecteur $\mathcal{C}^1$ défini sur un voisinage de $\gamma([0,1])$. On considère alors la carte

$$T_t(x) := \exp_x(t\xi(x)),$$

qui trace la géodésique partant de x et de vitesse $\xi(x)$. Pour avoir une idée de la courbure locale de notre espace, on veut calculer le Jacobien de cette application. Si on note $J(t)$ la matrice jacobienne de l'application T_t, dans une base orthonormale, les équations standards de la géométrie riemannienne donnent qu'elle doit vérifier l'équation de Jacobi suivante

$$J''(t) + R(t)J(t) = 0,$$

où R est une matrice symétrique construite à partir du tenseur de Riemann Riem_M de M, qui contient toutes les informations sur la métrique g_M. Alors, tant que $\det J(t) > 0$, on peut prendre la dérivée logarithmique de $\det(J)$, ce qui donne :

$$\frac{d}{dt}\log \det J(t) = \operatorname{tr}\left(J'(t)J(t)^{-1}\right).$$

Alors, en définissant $U(t) = J'(t)J(t)^{-1}$, l'équation de Jacobi implique que $U'(t) = -R(t) - U(t)^2$. Par suite, en prenant la trace de cette équation matricielle et en admettant que $\operatorname{tr} R(t) = \mathrm{Ric}_{\gamma(t)}$, on a :

$$\frac{d}{dt}\operatorname{tr} U(t) + \operatorname{tr} U(t)^2 + \mathrm{Ric}_{\gamma(t)}(\gamma'(t), \gamma'(t)) = 0.$$

Spécifions encore l'étude, et supposons que $\xi = \nabla\phi$ pour une certaine fonction ϕ. Cela implique que $U(0)$ est symétrique, et comme R est toujours symétrique et que U et $U^\star$ résolvent la même équation différentielle, $U(t)$ est également symétrique. Par l'inégalité de Cauchy-Schwarz, on alors que

$$\operatorname{tr} U(t)^2 \geq \frac{1}{n} \operatorname{tr}^2 U(t) \geq 0,$$

et d'où

$$\mathrm{Ric}_x \geq 0 \, \forall x \in M \iff \frac{d^2}{dt^2} \log \mathrm{Jac}_x(\exp_x(t\nabla\phi(x))) \leq 0 \, \forall \phi, \tag{ric}$$

où ϕ est arbitraire dans la classe des fonctions convexes laissant $\mathrm{Jac}_x(\exp_x(t\nabla\phi(x)))$ strictement positif sur $[0,1]$. Comme l'ont prouvé Otto et Villani dans les années 2000, la positivé de la courbure de Ricci peut se réexprimer très simplement en terme de convexité de l'entropie de Boltzmann H le long des géodésiques Wasserstein. On définit H de la façon suivante : soit $\mu \in \mathcal{P}(M)$, alors

$$H(\mu) := H_{\mathrm{vol}}(\mu) := \begin{cases} \displaystyle\int_M \rho \log \rho \, \mathrm{dvol} \text{ si } \mu = \rho \mathrm{vol}, \\ +\infty \text{ sinon.} \end{cases}$$

Comme nous l'évoquions dans la partie précédente, une des conséquences du théorème de McCann est que $(\mathcal{P}(M), W_2)$ est un espace géodésique, où

$$W_2(\mu,\nu) := \min_{\pi \in \Gamma(\mu,\nu)} \left\{ \int_{M\times M} d_M^2(x,y) \, \mathrm{d}\pi \right\}$$

est la distance de Wasserstein (ou Gromov-Wasserstein) d'ordre 2.

Théorème 2.2. *Soit (M,g) une variété riemannienne compacte. Alors $\mathrm{Ric} \geq 0$ si et seulement si $t \mapsto H(\mu_t)$ est une fonction convexe de $t \in [0,1]$ pour toutes les géodésiques Wasserstein $(\mu_t)_{0\leq t\leq 1}$.*

Preuve. (Idée de preuve du théorème d'Otto-Villani sur l'entropie de Boltzmann).

On prouve le sens direct seulement. Soit donc $(\mu_t)_{0\leq t\leq t}$ une géodésique Wasserstein. Un corollaire d'interpolation par déplacement du théorème de McCann nous donne que $\mu_t = T_{t\#}\mu_0$, où $T_t(x) = (1-t)x + tT(x)$ et $T = \nabla\phi$ est la carte de transport optimal de μ_0 à μ_1. Pour établir le théorème, on peut bien entendu se restreindre aux probabilités admettant une densité par rapport au volume. Formellement, un changement de variable donne

$$H(\mu_t) = \int_M \rho_t(x) \log(\rho_t(x)) \, \mathrm{dvol} = \int_M \rho_t(T_t(x)) \log(\rho_t(T_t(x))) \mathrm{Jac}_x T_t \, \mathrm{dvol}.$$

Mais par ailleurs, la condition $T_{t\#}\mu_0 = \mu_t$ impose qu'en termes de densités de probabilité, $\rho_t(T_t(x))\mathrm{Jac}_x(T_t(x)) = \rho_0(x)$: c'est l'équation de Monge-Ampère. La concavité de $t \mapsto \log \mathrm{Jac}_x(\exp_x(t\nabla\phi))$ donne alors le résultat :

$$H(\mu_t) = \int_M \rho_0(x) \log\left(\frac{\rho_0(x)}{\mathrm{Jac}_x T_t}\right) \mathrm{dvol} = H(\mu_0) - \int \log \mathrm{Jac}_x(\exp_x(t\nabla\phi)) \, \mathrm{d}\mu_0.$$

□

3 Focus sur un problème de flot de gradient

Pour une présentation plus détaillé du concept de flot des gradients, on pourra se référer au livre canonique [2].

3.1 Flot de gradient et espace de probabilité

Nous présentons dans cette partie des questions issues de l'article de MM. Carlier et Poon sur lequel j'ai pu travailler durant mon mémoire de M2, et que je vais probablement continuer à étudier pendant ma pré-thèse. En cherchant à modéliser le processus de débruitage d'une image, on obtient une équation non linéaire d'ordre 4 :

$$\partial_t \rho + \operatorname{div}\Big(\rho \nabla \operatorname{div}\Big(\frac{\nabla \rho}{|\nabla \rho|}\Big)\Big) = 0, \text{ dans } (0,\, T) \times \Omega,\ \rho_{t=0} = \rho_0, \tag{1}$$

avec Ω un ouvert borné. Il s'agit en fait d'une équation de diffusion très dégénérée, mais qu'on peut voir comme la limite d'un flot de gradient. Qu'est-ce qu'un flot de gradient ? Nous reprenons la présentation faite dans [13]. Très simplement, il s'agit de la discrétisation d'un problème de Cauchy

$$\begin{cases} x'(t) = -\nabla F(x(t)) \text{ pour } t > 0, \\ x(0) = x_0, \end{cases}$$

où l'on se donne $\tau > 0$ et x_k^τ et on cherche

$$x_{k+1}^\tau \in \operatorname{argmin}_x\ F(x) + \frac{|x - x_k^\tau|^2}{2\tau}.$$

Formellement, en x_{k+1}^τ, la dérivée de la fonction à minimiser s'annule et on obtient

$$\frac{x_{k+1}^\tau - x_k^\tau}{\tau} = -\nabla F(x_{k+1}^\tau),$$

ce qui est bien le schéma de résolution implicite du problème de Cauchy que l'on s'est donné. Pour $\tau \to 0$, on peut même s'attendre à ce que le schéma discret converge en un certain sens vers le schéma continu. Cette notion de convergence a été formalisée dans le concept de mouvement minimisant forgé par le mathématicien italien Ennio de Giorgi dans les années 1990 [8], qu'on peut utiliser dans les espaces de probabilité.

On définit J la fonctionnelle dite de Fokker-Planck qui agit sur les mesures de probabilité $\rho \in \mathcal{P}(K)$ via

$$J(\rho) = \int_\Omega \rho \log \rho + \int_\Omega V \,\mathrm{d}\rho,$$

où le terme entropique est infini si la probabilité n'est pas absolument continue par rapport à la mesure de Lebesgue. On définit alors le flot de gradient

$$\rho_{k+1}^\tau \in \operatorname{argmin}_\rho \Big\{ F(\rho) :=\ J(\rho) + \frac{W_2^2(\rho, \rho_k^\tau)}{2\tau} \Big\}, \tag{2}$$

dont une étude assez fine à l'aide d'outils d'analyse convexe permet de montrer qu'il converge vers notre équation d'ordre 4 1 quand $\tau \to 0$.

La question qui a lancé mes recherches de mémoire était celle du contrôle des solutions aux flots des gradients. Concrètement, si ρ_0^τ (on confond dans la suite les probabilités et leur densité) est la donnée initiale du flot de gradient et vérifie $0 < \alpha \leq \rho_0^\tau \leq \beta$ presque partout, la solution ρ_1^τ vérifie-t-elle le même encadrement ? Si cela était vrai, par transitivité, la solution limite ρ de 1 devrait aussi être encadrée, ce qui fournit des bornes a priori très intéressantes pour l'étude de l'équation en elle-même. Le principe du maximum a pu être démontré en faisant appel au concept de géodésiques généralisées :

Théorème 3.1. *Soit $\rho_0 \in \mathcal{P}_{\mathrm{ac}}(\Omega) \cap L^\infty(\Omega)$ et soit ρ_1 solution de 2. Alors $\rho_1 \in L^\infty(\Omega)$ avec*

$$\|\rho_1\|_{L^\infty(\Omega)} \leq \|\rho_0\|_{L^\infty(\Omega)}. \tag{3}$$

Preuve. (Idée de preuve du principe du maximum).

On se donne $\overline{\mu}, \mu_0$ et μ_1 dans $\mathcal{P}_{\mathrm{ac}}(\Omega)$, en notant T_0 (resp. T_1) la carte de Brenier entre $\overline{\mu}$ et μ_0 (resp. μ_1). La géodésique généralisée de base $\overline{\mu}$ joignant μ_0 to μ_1 est la courbe

$$\mu_t := ((1-t)T_0 + tT_1)_\# \overline{\mu}, \ t \in [0,1].$$

L'intérêt de cette définition de géodésiques est leur forte convexité

$$W_2^2(\overline{\mu}, \mu_t) \leq (1-t)W_2^2(\overline{\mu}, \mu_0) + tW_2^2(\overline{\mu}, \mu_1) - t(1-t)W_2^2(\mu_0, \mu_1),$$

et leur conservation des bornes supérieures

$$||\mu_t||_{L^p}^p \leq (1-t)||\mu_0||_{L^p}^p + t||\mu_1||_{L^p}^p, \ ||\mu_t||_{L^\infty} \leq \max(||\mu_0||_{L^\infty}, ||\mu_1||_{L^\infty}). \tag{4}$$

On définit alors $K := \{\rho \in \mathcal{P}_{\mathrm{ac}}(\Omega), \ \rho \leq ||\rho_0||_{L^\infty(\Omega)} \text{ a.e.}\}$. Grâce à l'inégalité précédente 4, on sait que les géodésiques généralisées joignant deux points de K restent dans K. Soit alors $\hat{\rho}_1$ la projection W_2 de ρ_1 sur K. Comme $\rho_t \in K$, on a $W_2^2(\rho_1, \hat{\rho}_1) \leq W_2^2(\rho_1, \rho_t)$. Alors par convexité forte appliquée à la géodésique généralisée de base ρ_1 joignant ρ_0 à $\hat{\rho}_1$, on a :

$$(1-t)W_2^2(\rho_1, \hat{\rho}_1) \leq (1-t)W_2^2(\rho_t, \rho_0) - t(1-t)W_2^2(\rho_0, \hat{\rho}_1),$$

soit $W_2^2(\rho_1, \hat{\rho}_1) \leq W^2(\rho_1, \rho_0) - W_2^2(\rho_0, \hat{\rho}_1)$. Une estimée subtile trouvée récemment par De Philipis et. al dans [12] donne que nécessairement $J(\hat{\rho}_1) \leq J(\rho_1)$. Alors, à moins que $W_2(\rho_1, \hat{\rho}_1) = 0$, $\hat{\rho}_1$ est meilleur que ρ_1 pour résoudre 2. D'où $\rho_1 = \hat{\rho}_1$, i.e. $\rho_1 \leq ||\rho_0||_{L^\infty}$.

□

3.2 Principe du minimum

La question de la stabilité des bornes inférieures des solutions du flot de Fokker-Planck s'avère plus difficile. En tout cas, il n'est plus possible d'utiliser des géodésiques généralisées, car on a un contrôle de leur borne supérieure mais pas de leur borne inférieure. En se plaçant en dimension un, on peut cependant exploiter le théorème de réarrangement monotone et la relation d'ordre de $\mathbb{R}$ pour obtenir le résultat.

Proposition 3.2. (Principe du minimum en dimension 1).

On suppose que $n = 1$, que Ω est un intervalle borné et que $\rho_0 \geq \alpha > 0$ presque partout sur Ω. Alors la solution ρ_1 de 2 vérifie également $\rho_1 \geq \alpha > 0$ p.p. sur Ω.

Pour démontrer le principe du minimum en toute généralité, une idée naturelle pourrait être de se ramener d'une manière ou d'une autre à la dimension 1, peut-être pourrons nous alors utiliser un réarrangement monotone. On a alors pensé a introduire la distance de Wasserstein slicée :

$$SW_2^2(\mu, \nu) = \fint_{\mathbb{S}^{d-1}} W^2(\theta_{\#}\mu, \theta_{\#}\nu)\,\mathrm{d}\theta,$$

où $\theta_{\#}\mu$ correspond au poussé en avant de μ par la projection orthogonale $\theta : x \mapsto \theta \cdot x$. La mesure $\theta_{\#}\mu$ est donc bien définie sur l'espace $\theta(\Omega)$ qui est de dimension 1. La question est alors de savoir si le flot de gradient SW

$$\rho_{k+1}^{\tau} \in \operatorname{argmin}_{\rho} \left\{ F(\rho) := \ J(\rho) + \frac{SW_2^2(\rho, \rho_k^{\tau})}{2\tau} \right\} \tag{5}$$

est équivalent au flot de gradient Wasserstein classique. Plusieurs résultats nous incitent à répondre par l'affirmative, comme l'équivalence des distances slicées et non slicées pour des mesures concentrées sur la boule de rayon R :

$$SW_2^2(\mu, \nu) \leq c_{d,2} W_2^2(\mu, \nu) \leq C_{d,2} R^{1/(d+1)} SW_2^{1/(d+1)}(\mu, \nu). \tag{6}$$

Pour poursuivre la comparaison de W et SW, nous nous sommes posés deux questions. La première portait sur le géodésisme éventuel de la distance SW. Il est bien connu que $(\mathcal{P}(\Omega), W_2)$ est un espace géodésique : pour $\mu, \nu \in \mathcal{P}(\Omega)$, on a

$$W_2(\mu, \nu) = \min \left\{ L_W(\omega),\, \omega \in AC(\mathcal{P}(\Omega), W_2),\, \omega(0) = \mu,\, \omega(1) = \nu \right\}.$$

Le terme $L_W(\omega)$ est la longueur de la courbe ω, supposée absolument continue, i.e. dont les variations sont contrôlées par une fonction L^1. Dans son livre [13], le mathématicien italien F. Santambrogio faisait la conjecture que $(\mathcal{P}(\Omega), SW_2)$ n'est au contraire pas un espace géodésique, mais un espace pseudo-géodésique dont la "vraie" distance géodésique est W_2. Il nous semble qu'à ce jour, nous avons démontré le résultat, à savoir que

$$\inf \left\{ L_{SW}(\omega),\, \omega \in AC(\mathcal{P}(\Omega), SW_2),\, \omega(0) = \mu,\, \omega(1) = \nu \right\} = c_d W_2(\mu, \nu).$$

La deuxième question qui nous intéresse est de savoir si le flot de gradient SW converge lui aussi vers l'équation 1 quand $\tau \to 0$. Dans l'hypothèse où nous arriverions à montrer que le flot de gradient SW vérifie le principe du minimum, nous aurions alors la stabilité des bornes portant sur les solutions de l'équation limite. Nous n'avons à ce jour pas encore résolu cette question.

Bibliographie

[1] L. Ambrosio. Lectures notes on optimal transport problem, 2003.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows : in metric spaces and in the space of probability measures.* Springer Science and Business Media, 2008.

[3] L. Ambrosio, D. Semola, and E. Brué. Lectures on optimal transport, 2019.

[4] Y. Brenier. Polar factorization and monotone rearrangement vector valued functions. *Comm. Pure Appl. Math.*, 44(4) :375–417, 1991.

[5] G. Carlier, A. Galichon, and F. Santambrogio. From knothe's transport to brenier's map and a continuation method for optimal transport. *SIAM J. Control Optim.*, 47(6) :2554–2576, 2010.

[6] A. Figalli, F. Maggi, and A. Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. *Invent math.*, 182 :167–211, 2010.

[7] A. Figalli and C. Villani. Lecture notes on optimal transport and curvature, 2008.

[8] E. De Giorgi. New problems on minimizing movements. *Boundary Value Problems for PDEs and Applications*, pages 81–98, 1993.

[9] L. V. Kantorovich. On a problem of monge. *Uspekhi Mat. Nauk.*, 3 :225–226, 1948. [English translation in J. Math. Sci. 133,4 (2006)].

[10] R. J. McCann. Polar factorization of maps on riemannian manifolds. *Geometric and Functional Analysis*, 11 :589–608, 2001.

[11] G. Monge. Mémoires de l'académie des sciences, 1781.

[12] G. De Philipis, R. Mésazàros, F. Santambrogio, and B. Velichkov. Bv estimates in optimal transportation and applications. *Archive for Rational Mechanincs and Analysis*, 219(2) :829–860, 2016.

[13] F. Santambrogio. *Optimal transport for applied mathematicians.* Birkhäuser/Springer, Cham, 2015.

[14] N. V. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.*, pages i–v, 1–178, 1979. [Cover to cover translation of Trudy Mat. Inst. Steklov vol. 141 (1976)].