Séminaire BOURBAKI Janvier 2013
65ème année, 2012-2013, n$^{\text{o}}$ 1065

FLOTS DE GRADIENT DANS LES ESPACES MÉTRIQUES ET LEURS APPLICATIONS [d'après Ambrosio–Gigli–Savaré]

par **Filippo SANTAMBROGIO**

INTRODUCTION

Pour traiter des flots de gradient dans les espaces métriques, il est sans doute plus judicieux de commencer cet exposé en disant de quoi il s'agit dans la situation la plus simple, c'est-à-dire dans l'espace euclidien. Étant donnés une fonction suffisamment régulière $F : \mathbb{R}^n \to \mathbb{R}$ et un point $x_0 \in \mathbb{R}^n$, un flot de gradient est une évolution, une courbe $x(t)$, dont le point de départ en $t = 0$ est x_0 et qui se déplace en choisissant à chaque instant la direction qui fait décroître F le plus rapidement possible. Plus précisément, ce n'est que la solution du *Problème de Cauchy*

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

Ce problème de Cauchy standard a une solution unique si ∇F est lipschitzien, c'est-à-dire si $F \in C^{1,1}$, mais on verra, entre autres, que l'existence et l'unicité pour cette solution pourront être obtenues sous des hypothèses beaucoup plus faibles, grâce à la structure variationnelle du problème.

À titre d'exemple, on peut voir que l'unicité est garantie dès que la fonction F est convexe, puisque pour deux solutions $x(t)$ et $y(t)$ on peut poser $E(t) = |x(t) - y(t)|^2$ et calculer

$$E'(t) = 2(x(t) - y(t)) \cdot (x'(t) - y'(t)) = -2(x(t) - y(t)) \cdot (\nabla F(x(t)) - \nabla F(y(t))) \leq 0,$$

l'inégalité venant de la propriété de monotonie $(\nabla F(x) - \nabla F(y)) \cdot (x - y) \geq 0$, vérifiée pour tous $x, y \in \mathbb{R}^n$ dès lors que F est convexe. Ceci entraîne évidemment l'égalité $x(t) = y(t)$ si $x(0) = y(0)$, et donne aussi une estimation de stabilité, la distance entre deux solutions au temps t s'estimant avec celle entre les données initiales.

De même, si F n'est pas convexe mais juste λ-convexe (c'est-à-dire que $x \mapsto F(x) + \lambda\frac{|x|^2}{2}$ est convexe, une condition plus forte que la convexité si $\lambda > 0$ et plus faible si $\lambda < 0$), on peut également conclure à une estimation de ce genre grâce au Lemme de Gronwall. En effet, dans ce cas on a $(\nabla F(x) - \nabla F(y)) \cdot (x - y) \geq \lambda|x - y|^2$, et donc

$$E'(t) \leq -2\lambda E(t) \Rightarrow E(t) \leq e^{-2\lambda t}E(0),$$

ce qui entraîne encore l'unicité et la stabilité. Le fait de savoir traiter des fonctions λ-convexes (c'est-à-dire des fonctions dont les dérivées secondes sont bornées inférieurement) est un point important, en particulier parce que cela fait une extension du cas $C^{1,1}$ (toute fonction $C^{1,1}$ étant λ-convexe pour un certain λ).

Or, notre but principal est d'étendre cette théorie au cas où l'espace euclidien $\mathbb{R}^n$ est remplacé par un espace métrique. Cela n'est pas du tout évident et nécessite de donner les définitions opportunes, qui n'utilisent pas la notion de gradient ∇F. Aussi, il va sans dire que, si on sait le faire, on sait également se passer de l'hypothèse $F \in C^{1,1}$.

Une première interprétation des flots de gradient qui ne requiert pas la structure différentielle vient de leur discrétisation en temps. En effet, si l'on fixe un pas de temps $\tau > 0$, on peut considérer, au lieu de l'équation différentielle, la suite $(x_k^\tau)_k$ définie par récurrence de la manière suivante :

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

(indépendamment du fait que le problème de minimisation ci-dessus admette ou pas une solution unique). Ce qui est important de cette suite est qu'on peut interpréter ses points comme les valeurs de la courbe $x(t)$ aux instants $t = 0, \tau, 2\tau, \dots, k\tau, \dots$. En effet, les conditions d'optimalité de cette suite récursive de problèmes d'optimisation nous donnent exactement

$$x_{k+1}^\tau \in \operatorname{argmin} F(x) + \frac{|x - x_k^\tau|^2}{2\tau} \Rightarrow \nabla F(x_{k+1}^\tau) + \frac{x_{k+1}^\tau - x_k^\tau}{\tau} = 0,$$
$$\text{i.e.} \quad \frac{x_{k+1}^\tau - x_k^\tau}{\tau} = -\nabla F(x_{k+1}^\tau).$$

Cette expression correspond à ce que l'on appelle *schéma d'Euler implicite* de l'équation $x' = -\nabla F(x)$. Si on prouve que, pour $\tau \to 0$, la suite qu'on a trouvée, interpolée de manière opportune, converge à la solution du problème, alors on soupçonne qu'on pourrait même définir une notion de flot de gradient pour une fonction F qui satisfasse juste les hypothèses aptes à donner l'existence d'un minimiseur à chaque étape (F semi-continue inférieurement et quelques hypothèses de compacité).

Encore mieux, on s'aperçoit que cette formulation discrétisée en temps peut s'adapter parfaitement au cas où $\mathbb{R}^n$ est remplacé par un espace métrique. Si l'on a un espace métrique (X, d) (compact, par exemple) et une fonction $F : X \to \mathbb{R} \cup \{+\infty\}$ semi-continue, on peut définir la suite

$$x_{k+1}^\tau \in \operatorname{argmin}_x F(x) + \frac{d(x, x_k^\tau)^2}{2\tau}, \tag{1}$$

l'interpoler de manière constante par morceaux

$$x^\tau(t) := x_k^\tau \quad \text{pour tout } t \in [k\tau, (k+1)\tau[\tag{2}$$

et étudier les limites des courbes x^τ lorsque $\tau \to 0$.

De Giorgi, dans [DeG], définissait ainsi les *mouvements minimisant généralisés* :

Définition 0.1. — *Une courbe $x : [0,T] \to X$ est dite Mouvement Minimisant Généralisé (MMG) si il existe une suite de pas de temps $\tau_j \to 0$ telle que la suite de courbes x^{τ_j} définies en* (2) *en partant d'une suite de solutions du schéma discret* (1) *converge uniformément à x sur $[0,T]$.*

Les résultats de compacité garantissant l'existence d'un tel mouvement minimisant généralisé découlent d'une propriété assez simple de presque-continuité Hölder : pour tout τ et tout k, l'optimalité de x^τ_{k+1} nous donne

$$F(x^\tau_{k+1}) + \frac{d(x^\tau_{k+1}, x^\tau_k)^2}{2\tau} \leq F(x^\tau_k), \tag{3}$$

ce qui entraîne

$$d(x^\tau_{k+1}, x^\tau_k)^2 \leq 2\tau \left(F(x^\tau_k) - F(x^\tau_{k+1})\right).$$

Si $F(x_0)$ est fini et F est bornée inférieurement, en prenant la somme sur k on a

$$\sum_{k=0}^{l} d(x^\tau_{k+1}, x^\tau_k)^2 \leq 2\tau \left(F(x^\tau_0) - F(x^\tau_{l+1})\right) \leq C\tau.$$

L'inégalité de Cauchy-Schwartz nous donne alors, si $t < s$, $t \in [k\tau, (k+1)\tau[$ et $s \in [l\tau, (l+1)\tau[$ (et donc $|l-k| \leq \frac{|t-s|}{\tau} + 1$)

$$d(x^\tau(t), x^\tau(s)) \leq \sum_{k=0}^{l} d(x^\tau_{k+1}, x^\tau_k) \leq \left(\sum_{k=0}^{l} d(x^\tau_{k+1}, x^\tau_k)^2\right)^{1/2} \left(\frac{|t-s|}{\tau} + 1\right)^{1/2} \leq C\left(|t-s|^{1/2} + \sqrt{\tau}\right).$$

Cela nous dit que les courbes x^τ – si on oublie qu'elles sont discontinues – sont moralement équi-höldériennes d'exposant 1/2, et permet d'extraire une sous-suite convergente.

Or, si l'espace X, la distance d, et la fonctionnelle F sont connus explicitement, dans certains cas il est déjà possible de passer à la limite dans les conditions d'optimalité de chaque problème d'optimisation en temps discret, et de caractériser les courbes (ou la courbe) limite $x(t)$. Il sera possible de faire ainsi dans le cadre des mesures de probabilité dont il est question dans la section 2, mais pas dans d'autres cas. De fait, sans un petit peu de structure (différentielle) sur l'espace X, cela est pratiquement impossible. Si l'on souhaite développer une théorie générale pour les flots de gradient dans les espaces métriques, il faut utiliser des instruments plus fins, qui permettent vraiment de caractériser, à l'aide seulement de quantités métriques, le fait qu'une courbe continue $x(t)$ soit un flot de gradient. Le livre d'Ambrosio-Gigli-Savaré [AGS05], et en particulier sa première partie (la deuxième étant dédiée aux espaces de mesures de probabilité), se donne exactement cet objectif.

Nous présentons ici deux inégalités qui sont satisfaites par les flots de gradient dans le cas euclidien régulier, et qui peuvent être utilisées comme définition de flot de gradient dans un cadre métrique, toutes les quantités qui y apparaissent ayant une contrepartie métrique.

La première observation est la suivante : pour toute courbe $x(t)$ on a

$$F(x(s)) - F(x(t)) = \int_s^t -\nabla F(x(r)) \cdot x'(r)\, dr \le \int_s^t |\nabla F(x(r))||x'(r)|\, dr$$
$$\le \int_s^t \left(\frac{1}{2}|x'(r)|^2 + \frac{1}{2}|\nabla F(x(r))|^2 \right) dr.$$

Ci-dessus, la première inégalité est une égalité si et seulement si $x'(r)$ et $\nabla F(x(r))$ sont des vecteurs de directions opposées pour presque tout r, et la deuxième est une égalité si et seulement si leurs modules sont égaux. Ainsi, la condition, appelée EDE (*Energy Dissipation Equality*)

$$F(x(s)) - F(x(t)) = \int_s^t \left(\frac{1}{2}|x'(r)|^2 + \frac{1}{2}|\nabla F(x(r))|^2 \right) dr, \quad \text{pour tous } s < t$$

(ou même la simple inégalité $F(x(s)) - F(x(t)) \ge \int_s^t \left(\frac{1}{2}|x'(r)|^2 + \frac{1}{2}|\nabla F(x(r))|^2\right) dr$) est équivalente à $x' = -\nabla F(x)$ p.p., et pourrait être prise comme définition de flot de gradient. On verra que les deux objets $|x'|$ et $|\nabla F|$ ont un sens dans les espaces métriques, et que cela donne des résultats très puissants d'existence.

Pour les résultats d'unicité, une autre caractérisation est proposée. Elle se base sur l'observation suivante : si $F : \mathbb{R}^n \to \mathbb{R}$ est convexe, alors l'inégalité

$$F(y) \ge F(x) + p \cdot (y - x) \quad \text{pour tout } y \in \mathbb{R}^n$$

caractérise (par définition) les vecteurs $p \in \partial F(x)$ et, si $F \in C^1$, elle est vérifiée uniquement par $p = \nabla F(x)$. De même, si F est λ-convexe, l'inégalité qui caractérise le gradient est

$$F(y) \ge F(x) + \frac{\lambda}{2}|x - y|^2 + p \cdot (y - x) \quad \text{pour tout } y \in \mathbb{R}^n.$$

Ainsi, on peut prendre une courbe $x(t)$ et un point y et calculer

$$\frac{d}{dt}\frac{1}{2}|x(t) - y|^2 = (y - x(t)) \cdot (-x'(t)).$$

Par conséquent, imposer que, pour tout t et tout y, on ait

$$\frac{d}{dt}\frac{1}{2}|x(t) - y|^2 \le F(y) - F(x(t)) - \frac{\lambda}{2}|x(t) - y|^2,$$

sera donc équivalent à l'égalité $-x'(t) = \nabla F(x(t))$ pour tout t. Cela donnera une deuxième caractérisation (appelée EVI, *Evolution Variational Inequality*) des flots de gradient dans un environnement métrique. Même si on oubliera souvent la dépendance en λ, il faut remarquer que la condition EVI devrait être indiquée comme EVI_λ, puisqu'elle fait intervenir un paramètre λ, a priori arbitraire. D'ailleurs, remarquons aussi que la λ-convexité de F n'est pas nécessaire pour définir la propriété EVI_λ, mais le sera pour l'existence de courbes la satisfaisant, ce qui nous amènera à définir ce qu'est une fonction λ-convexe dans un espace métrique (ce sera d'ailleurs une propriété nécessaire pour l'existence de ces courbes).

Plan de l'exposé. Jusque-là on a vu comment certaines notions dans la théorie euclidienne des flots de gradient pourraient s'exprimer à l'aide de quantités métriques : dans la suite – structurée bien entendu en trois parties, chacune composée de trois sous-parties – on verra d'abord par quelles techniques on pourra avoir des résultats d'existence et unicité pour les flots de gradient (définis à l'aide des conditions EDE ou EVI) dans un espace métrique (Section 1) ; ensuite viendra le tour du cas particulier de l'espace des mesures de probabilité muni d'une distance issue du transport optimal et des EDP d'évolution associées à ces flots de gradient (Section 2) ; on terminera par des résultats récents de Gigli et de ses collaborateurs, motivés par l'observation que le flot de la chaleur est en même temps un flot de gradient par rapport à la métrique du transport et par rapport à la distance L^2 (Section 3).

1. LA THÉORIE GÉNÉRALE DANS LES ESPACES MÉTRIQUES

1.1. Préliminaires métriques

Pour esquisser une théorie générale dans les espaces métriques, il est tout d'abord nécessaire de définir au moins les trois objets dont on a besoin pour parler des propriétés EDE et EVI caractérisant les flots de gradient : la notion de vitesse d'une courbe, celle de pente d'une fonction, et celle de convexité géodésique.

Dérivée métrique. Pour toute courbe $x : [0, T] \to X$ à valeur dans un espace métrique on peut définir, au lieu de la vitesse $x'(t)$ en tant que vecteur (avec sa direction, comme on le ferait dans un espace vectoriel), le module de sa vitesse :

$$|x'|(t) := \lim_{h\to 0} \frac{d(x(t), x(t+h))}{|h|},$$

à condition que la limite existe. On voit facilement que cette définition coïncide avec la norme de la dérivée si l'on est dans un espace normé. Dans l'esprit du Théorème de Rademacher, on peut démontrer que cette limite existe pour presque tout t si la courbe x est lipschitzienne (il est facile de le faire dans l'espace métrique l^∞, en prenant le sup des modules des vitesses des composantes, et la généralisation à tout espace métrique compact se fait en le plongeant dans l^∞ ; la preuve peut être restreinte au cas des espaces compacts, puisque l'image de $[0, T]$ par la courbe continue x l'est toujours). De même, par reparamétrage, cela s'étend aux courbes absolument continues.
Celle-ci est la notion de vitesse qu'on utilisera dans les espaces métriques. On remarque que la longueur d'une courbe est bien l'intégrale de sa dérivée métrique, et qu'on peut reparamétrer les courbes de longueur finie de manière à ce que leur dérivée métrique soit constante.

Pente et module du gradient. Plusieurs définitions différentes du module du gradient d'une fonction F définie sur un espace métrique sont possibles. Tout d'abord on peut appeler *gradient supérieur* (*upper gradient*, en anglais) toute fonction $g : X \to \mathbb{R}$ telle que, pour toute courbe lipschitzienne x, on a

$$|F(x(0)) - F(x(1))| \leq \int_0^1 g(x(t))|x'|(t)dt.$$

Si F est lipschitzienne, un choix possible est la *constante de Lipschitz locale*

$$|\nabla F|(x) := \limsup_{y \to x} \frac{|F(x) - F(y)|}{d(x,y)}; \tag{4}$$

un autre est la pente descendante (*descending slope* en anglais, mais on l'indiquera simplement *pente* en français), qui est une notion en soi plus adaptée à la minimisation d'une fonction qu'à sa maximisation, et donc a priori raisonnable pour des fonctions semi-continues inférieurement :

$$|\nabla^- F|(x) := \limsup_{y \to x} \frac{[F(x) - F(y)]_+}{d(x,y)}$$

(remarquons que la pente est nulle en tout point de minimum local). En général il n'est pas vrai que la pente soit un gradient supérieur, même si on donnera des cas où elle l'est. On verra ensuite (Section 3) comment, pour établir une notion d'espace H^1 sur un espace métrique, on prendra des notions « relaxées » du module du gradient de F.

Convexité géodésique. Troisième notion à fixer, celle de convexité. Cela ne marche que dans un espace géodésique, c'est-à-dire dans un espace métrique (X, d) tel que, pour tout couple $(x(0), x(1))$ de points de X, il existe une courbe $x : [0,1] \to X$ les connectant et telle que $d(x(t), x(s)) = |t - s| d(x(0), x(1))$. Une telle courbe est forcément une *géodésique* (courbe de longueur minimale connectant deux points donnés) et sa vitesse $|x'|(t)$ est constante et égale à $d(x(0), x(1))$. On peut alors définir les fonctions *géodésiquement convexes* comme ces fonctions $F : X \to \mathbb{R} \cup \{+\infty\}$ qui sont convexes le long des géodésiques. Plus précisément, on demande à ce que pour tout couple $(x(0), x(1))$ il existe [1] une géodésique x à vitesse constante connectant les deux points telle que

$$F(x(t)) \leq (1-t)F(x(0)) + tF(x(1)).$$

On peut également définir les fonctions λ-convexes comme celles qui satisfont une version modifiée de l'inégalité ci-dessus, à savoir

$$F(x(t)) \leq (1-t)F(x(0)) + tF(x(1)) - \lambda \frac{t(1-t)}{2} d^2(x(0), x(1)).$$

1. Attention, cette définition n'équivaut pas à la vraie convexité le long de la géodésique, parce qu'on ne compare que les instants intermédiaires t aux instants 0 et 1, pas entre eux ; toutefois, en cas d'unicité de la géodésique ou si on demandait à ce que cette condition soit valable pour toute géodésique cela reviendrait au même. Aussi, on peut remarquer qu'on n'aurait besoin de l'existence de géodésiques que pour connecter les points où $F < +\infty$.

1.2. Existence d'un flot de gradient

Une fois établies ces notions de bases, on passe maintenant aux notions de flots de gradient. L'approche est encore et toujours celle de procéder à une minimisation par étapes temporelles pour un pas de temps $\tau > 0$ fixé, et de passer à la limite ensuite. On veut tout d'abord préciser le cadre dans lequel cela est bien posé. On suppose que l'espace X et la fonction F sont tels que tout ensemble de sous-niveau $\{F \leq c\}$ soit compact dans X, soit pour la topologie donnée par la distance d, soit pour une topologie plus faible, par rapport à laquelle la distance d est semi-continue inférieurement ; F serait aussi semi-continue par rapport à la même topologie. Ceci est le cadre minimal qui garantit l'existence des minimiseurs à chaque étape et le fait que des estimations du type de (3) donnent l'existence d'une courbe limite. C'est d'ailleurs une situation assez générale, comme on peut le voir dans le cas où X est un espace de Banach réflexif et la distance celle induite par la norme : se limiter aux fonctions F fortement continues serait trop contraignant, mais la topologie faible fait typiquement l'affaire.

On comprend aisément que, bien que l'estimation (3) soit suffisante à donner la compacité, et donc l'existence d'un MMG, elle ne pourra jamais caractériser la courbe limite (elle est satisfaite par toute évolution discrète où x^τ_{k+1} donne une valeur meilleure que x^τ_k, sans aucune optimalité requise) et donc on n'obtiendra a priori aucune des deux formulations – EDE ou EVI – des flots de gradient métriques.

Pour améliorer le résultat, il faut mieux exploiter « combien » x^τ_{k+1} fait mieux que x^τ_k. Une idée due à De Giorgi permet d'obtenir le résultat souhaité, et ce par une interpolation « variationnelle » entre les points x^τ_k et x^τ_{k+1}. Pour ce faire, x^τ_k étant fixé pour toute valeur du paramètre $\theta \in]0,1]$, on peut considérer

$$\min_x \quad F(x) + \frac{d^2(x, x^\tau_k)}{2\theta\tau}$$

et appeler $x(\theta)$ un minimiseur de ce problème, et $v(\theta)$ la valeur minimale. Il est clair que, pour $\theta \to 0^+$, on a $x(\theta) \to x^\tau_k$ et $v(\theta) \to F(x^\tau_k)$, et que, pour $\theta = 1$, on revient au problème et au minimiseur x^τ_{k+1} usuels. De plus, la fonction v est monotone décroissante et, par conséquent, dérivable presque partout (on peut même démontrer qu'elle est localement semi-concave). Sa dérivée $v'(\theta)$ est donnée par la dérivée de la fonction $\theta \mapsto F(x) + \frac{d^2(x,x^\tau_k)}{2\theta\tau}$, calculée au point $x = x(\theta)$ optimal (l'existence de $v'(\theta)$ implique que cette dérivée est la même quel que soit le minimiseur $x(\theta)$). On a donc

$$v'(\theta) = -\frac{d^2(x(\theta), x^\tau_k)}{2\theta^2\tau},$$

ce qui implique d'ailleurs que la distance $d(x(\theta), x^\tau_k)^2$ ne dépend pas du minimiseur $x(\theta)$. Aussi, les conditions d'optimalité pour le problème de minimisation à $\theta > 0$ fixé nous montrent assez facilement que $|\nabla^- F|(x(\theta)) \leq d(x(\theta), x^\tau_k)/\theta\tau$. Ceci peut se voir si on considère la minimisation d'une fonction $x \mapsto F(x) + cd^2(x, \bar{x})$, pour $c > 0$ et $\bar{x}$ quelconques, et on considère un compétiteur y. Si x est optimal on a

$$\begin{aligned} &F(y) + cd^2(y, \bar{x}) \geq F(x) + cd^2(x, \bar{x}) \\ &\quad \Rightarrow F(x) - F(y) \leq c \left(d^2(y, \bar{x}) - d^2(x, \bar{x})\right) = c \left(d(y, \bar{x}) + d(y, \bar{x})\right) \left(d(y, \bar{x}) - d(y, \bar{x})\right) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \leq c \left(d(y, \bar{x}) + d(y, \bar{x})\right) d(y, x). \end{aligned}$$

En divisant par $d(y, x)$ et en prenant la lim sup pour $y \to x$, on a $|\nabla^- F|(x) \leq 2cd(x, \bar{x})$.

On revient à la fonction v et on utilise $v(0) - v(1) = -\int_0^1 v'(\theta)d\theta$, avec l'inégalité

$$-v'(\theta) = \frac{d(x(\theta), x_k^\tau)^2}{2\theta^2 \tau} \geq \frac{\tau}{2} |\nabla^- F(x(\theta))|^2$$

que l'on vient d'obtenir. Il en réxulte donc une version améliorée de (3) :

$$F(x_{k+1}^\tau) + \frac{d(x_{k+1}^\tau, x_k^\tau)^2}{2\tau} \leq F(x_k^\tau) - \frac{\tau}{2} \int_0^1 |\nabla^- F(x(\theta))|^2 d\theta.$$

Si on ajoute ces inégalités pour $k = 0, 1, 2, \ldots$ et qu'ensuite on passe à la limite $\tau \to 0$, on peut obtenir, pour tout MMG x, l'inégalité

$$F(x(t)) + \frac{1}{2} \int_0^t |x'|(r)^2 dr + \frac{1}{2} \int_0^t |\nabla^- F(x(r))|^2 dr \leq F(x(0)), \tag{5}$$

pourvu que des hypothèses opportunes aient été établies. En particulier, il faut la semi-continuité inférieure de F pour gérer le terme $F(x_{k+1}^\tau)$ qui donnera $F(x(t))$, mais il faut également la semi-continuité de la pente $|\nabla^- F|$ pour gérer le terme correspondant.

Cette inégalité ne correspond pas exactement à EDE : d'une part on a une inégalité, et d'autre part on compare les instants t et 0 au lieu de t et s. Si on veut obtenir cette propriété plus forte on peut demander à ce que la pente soit un gradient supérieur. Ceci parce que, dans ce cas, on a l'inégalité $F(x(0)) - F(x(t)) \leq \int_0^t |\nabla^- F(x(r))||x'|(r)dr$ et, en continuant avec les inégalités usuelles, on trouve que (5) est en fait une égalité. Cela permet de soustraire l'égalité pour s de celle pour t, et obtenir, pour tout $s < t$:

$$F(x(t)) + \frac{1}{2} \int_s^t |x'|(r)^2 dr + \frac{1}{2} \int_s^t |\nabla^- F(x(r))|^2 dr = F(x(s)).$$

Or, il se trouve que si on suppose que F est λ-géodésiquement convexe, le cadre se simplifie fortement. En effet, on a deux avantages : la semi-continuité de la pente, ainsi que le fait qu'elle est un gradient supérieur. Ces deux résultats sont prouvés dans [AGS05, AG]. Nous ne donnons ici que l'idée principale qui permet de prouver les deux. Elle repose sur une représentation ponctuelle de la pente comme un sup au lieu d'une lim sup : si F est λ-géodésiquement convexe, en effet, on peut vérifier qu'on a

$$|\nabla^- F|(x) = \sup_{y \neq x} \left[\frac{F(x) - F(y)}{d(x, y)} + \frac{\lambda}{2} d(x, y) \right]_+ . \tag{6}$$

Pour vérifier cette égalité, il suffit de rajouter un terme $\frac{\lambda}{2} d(x, y)$ à l'intérieur de la partie positive de la définition de $|\nabla^- F|(x)$, ce qui ne change pas la limite $y \to x$ et montre que $|\nabla^- F|(x)$ est inférieure ou égale à ce sup, et prouver l'inégalité opposée en fixant un y, le connectant à x par une géodésique $x(t)$, et en calculant ensuite la limite le long de cette géodésique.

Cette représentation comme un sup permet aussi de démontrer la semi-continuité de la pente (2). Il est également possible (voir [AG], par exemple) de prouver que la pente est un gradient supérieur.

J'insiste néanmoins sur le fait que l'hypothèse de λ-convexité n'est pas naturelle ni cruciale pour l'existence d'un flot de gradient. D'une part, parce que des fonctions suffisamment régulières pourraient satisfaire également les hypothèses de semi-continuité de F et de la pente $|\nabla^- F|$ et le fait que $|\nabla^- F|$ soit un gradient supérieur indépendamment de la convexité, et d'autre part parce que le schéma discret donne déjà une méthode, bien définie sous des hypothèses bien plus faibles, pour trouver une courbe limite. Si l'espace et la fonctionnelle le permettent (comme ce sera le cas dans la prochaine section) on peut espérer caractériser cette courbe comme la solution d'une équation (aux dérivées partielles dans la section 2) sans passer par la théorie générale et par la condition EDE.

1.3. Unicité et contractivité

Au contraire, au vu de la preuve d'unicité qu'on avait donnée dans le cas euclidien dans l'introduction, une certaine convexité semble être une condition à demander si on veut des résultats d'unicité.

Pour ne pas alourdir l'exposition, on se limitera ici à donner les lignes générales de ce qui est la théorie de l'unicité dans un cadre métrique. Le point clé est qu'elle se base sur la condition EVI au lieu de la condition EDE. La relation entre ces deux conditions a été étudiée et clarifiée par Savaré (dans un papier non publié, mais la preuve peut se trouver dans [AG]), qui a montré que

- Toute courbe qui est un flot de gradient au sens de la condition EVI l'est aussi au sens de la condition EDE.
- La condition EDE n'est en général pas suffisante pour garantir l'unicité d'une solution (3). En revanche, il y a pratiquement toujours existence d'un flot de gradient au sens EDE.
- La condition EVI est en général trop contraignante pour garantir l'existence d'un flot de gradient en ce sens, mais elle permet toujours d'établir l'unicité et même la stabilité (par rapport aux données initiales) des flots de gradient.

Remarquons au passage que l'existence d'un flot de gradient au sens EVI est très contraignante pour la fonction (on verra ensuite qu'elle l'est aussi pour l'espace), puisqu'un résultat contenu dans [DS] affirme que si F est telle que, pour tout point de

2. Attention, dans ce cas on a semi-continuité par rapport à la topologie de la distance d, ce qui permet de gérer seulement le cas où les ensembles de niveau $\{F \leq c\}$ sont d-compacts.

3. L'exemple le plus simple est le suivant : prenons $X = \mathbb{R}^2$ muni de la distance l^∞ donnée par $d((x_1, x_2), (y_1, y_2)) = |x_1 - y_1| \vee |x_2 - y_2|$, et $F(x_1, x_2) = x_1$; il se trouve que toute courbe $(x_1(t), x_2(t))$ avec $x_1'(t) = -1$ et $|x_2'(t)| \leq 1$ satisfait la condition EDE ; dans le même exemple, on peut également voir qu'il n'y a pas de courbes satisfaisant EVI.

départ x_0, il existe un flot de gradient satisfaisant la condition EVI$_\lambda$, alors F est forcément λ-géodésiquement convexe.

Nous donnons une idée de preuve de la contractivité (et donc de l'unicité) des flots de gradient EVI :

Proposition 1.1. — *Si deux courbes x et $y : [0,T] \to X$ satisfont la condition EVI, alors on a*

$$\frac{d}{dt} d(x(t), y(t))^2 \le -2\lambda d(x(t), y(t))^2$$

et $d(x(t), y(t)) \le e^{-\lambda t} d(x(0), y(0))$.

La deuxième partie de l'énoncé est une conséquence de la première par le Lemme de Gronwall. La première est obtenue (formellement) en dérivant $t \mapsto d(x(t), y(t_0))^2$ par rapport à t en $t = t_0$, puis $s \mapsto d(x(t_0), y(s))^2$ par rapport à s, en $s = t_0$. L'inégalité qui définit la condition EVI permet de dire

$$\frac{d}{dt} d(x(t), y(t_0))^2_{|t=t_0} \le -\lambda d(x(t_0), y(t_0))^2 + 2F(y(t_0)) - 2F(x(t_0))$$
$$\frac{d}{ds} d(x(t_0), y(s))^2_{|s=t_0} \le -\lambda d(x(t_0), y(t_0))^2 + 2F(x(t_0)) - 2F(y(t_0))$$

et donc, en ajoutant les deux inégalités, et en jouant avec les dérivées composées,

$$\frac{d}{dt} d(x(t), y(t))^2 \le -2\lambda d(x(t), y(t))^2.$$

Si l'on veut une théorie satisfaisante de l'unicité pour les flots de gradient, il reste donc à démontrer l'existence de courbes qui satisfassent la condition EVI, sachant que cela demandera sans doute des hypothèses additionnelles. Cela se fait encore grâce au schéma discret, en rajoutant une hypothèse de compatibilité entre la fonction F et la distance d, condition qui fait intervenir une certaine notion de convexité. Nous ne rentrons pas dans les détails de la preuve, pour laquelle on renvoie à [AGS05], où est prouvée la convergence vers un flot de gradient EVI, avec des estimations d'erreur explicites. Ces estimations a priori permettent d'ailleurs de démontrer qu'on a affaire à une suite de Cauchy, en se passant donc de la compacité pour avoir la convergence (on pourrait même se passer de l'hypothèse de compacité dont on aurait besoin pour démontrer l'existence d'une solution à chaque étape temporelle discrète, si on utilise de presque-minimiseurs en faisant appel au principe variationnel d'Ekeland). Nous nous limitons à discuter cette hypothèse ultérieure de convexité.

L'hypothèse, qu'on appellera C^{2}G^2 (*Compatibilité du Comportement sur les Géodésiques Généralisées*) est la suivante : supposons que, pour toute paire (x_0, x_1) et tout point $y \in X$, il existe une courbe $x(t)$ connectant $x(0)$ à $x(1)$, telle que

$$F(x(t)) \le (1-t)F(x(0)) + tF(x(1)) - \lambda \frac{t(1-t)}{2} d^2(x(0), x(1)),$$
$$d^2(x(t), y) \le (1-t)d^2(x(0), y) + td^2(x(1), y) - t(1-t)d^2(x(0), x(1)).$$

Autrement dit, on demande la λ-convexité de la fonction F mais également la 2-convexité de la fonction $x \mapsto d^2(x,y)$, le long d'une même courbe qui n'est pas forcément la géodésique. Il faut remarquer que cette deuxième condition est automatiquement vérifiée, en prenant la géodésique, dans l'espace euclidien (et dans tout espace de Hilbert) puisque la fonction $x \mapsto |x-y|^2$ est quadratique avec matrice hessienne égale à $2I$ en tout point. On peut également vérifier qu'elle n'est satisfaite dans un espace normé que si la norme provient d'un produit scalaire. Il a été remarqué récemment par Gigli que la bonne condition pour avoir l'existence des flots de gradient au sens EVI est en effet que l'espace soit « infinitésimalement hilbertien » (en un sens à préciser, on y reviendra à la fin de la section 3).

Ici, on se contente de remarquer que la $\mathrm{C}^2\mathrm{G}^2$ implique la $(\lambda+\frac{1}{\tau})$-convexité, le long de ces courbes (parfois appelées *géodésiques généralisées*, sachant que ce sont des courbes qui dépendent d'un point base typiquement extérieur aux deux points à connecter), de la fonctionnelle qu'on minimise à chaque étape du schéma discret. Cela donne l'unicité du minimiseur pour τ petit, et permet de faire les estimations voulues.

Aussi, le choix de cette hypothèse a été fait en vue des applications aux espaces de Wasserstein, qu'on présentera dans la prochaine section. En effet, dans ces espaces le carré de la distance n'est pas en général 2-convexe mais on peut trouver, pour beaucoup de fonctionnelles F, des courbes adaptées qui permettent d'avoir la convexité opportune.

2. LES FLOTS DE GRADIENT DANS L'ESPACE DE WASSERSTEIN

L'une des applications les plus excitantes (et peut-être la seule [4] véritable application, au sens des mathématiques appliquées) de la théorie des flots de gradient dans les espaces métriques est sans doute celle aux EDP d'évolutions dans des espaces de mesures. Ce sujet puise son inspiration dans les travaux de Jordan, Kinderlehrer et Otto ([JKO]), qui avaient eu l'intuition que l'équation de la chaleur et celle de Fokker-Planck ont une structure variationnelle commune vis-à-vis d'une métrique particulière sur les mesures de probabilité, la distance de Wasserstein. Cependant, ce n'est qu'avec les travaux d'Ambrosio, Gigli et Savaré que la théorie a été formalisée dans un cadre plus abstrait et général (sans que cela signifie que la preuve de [JKO] n'était pas rigoureuse, c'est l'intuition sur la structure générale qui nécessitait encore une mise au point).

L'idée principale est celle de munir l'espace $\mathcal{P}(\Omega)$ des mesures de probabilité sur un domaine Ω d'une distance et de traiter ensuite les flots de gradient de certaines fonctionnelles sur cet espace métrique.

Pour expliquer de quoi il s'agit, nous allons d'abord donner un bref aperçu de cette distance, issue de la théorie du transport optimal. On peut trouver plus de détails sur

4. Qu'il s'agisse de la seule est sûrement exagéré ; on pourrait penser par exemple à la théorie des évolutions géométriques de formes et d'ensembles, même s'il semble que cette approche métrique ne se soit pas encore imposée.

cette théorie dans les ouvrages de C. Villani ([Vil03, Vil09]), ainsi que dans le livre [AGS05] (5).

2.1. Transport optimal et distance W_2

La distance sur les mesures de probabilité qu'on va considérer est celle, désormais appelée *Distance de Wasserstein* (6), issue du problème de transport optimal de Monge-Kantorovitch. Étant données deux mesures de probabilité $\mu, \nu \in \mathcal{P}(\Omega)$ sur un domaine $\Omega \subset \mathbb{R}^d$ (qu'on peut prendre compact par simplicité d'exposition), le problème du transport optimal consiste en

$$\min\left\{\int_{\Omega\times\Omega}|x-y|^2\,d\gamma\;:\;\gamma\in\Pi(\mu,\nu)\right\},$$

où $\Pi(\mu,\nu)$ est l'ensemble des *plans de transport*, c'est-à-dire

$$\Pi(\mu,\nu)=\{\gamma\in\mathcal{P}(\Omega\times\Omega):\,(p^x)_\#\gamma=\mu,\;(p^y)_\#\gamma=\nu,\},$$

p^x et p^y étant les deux projections d'$\Omega\times\Omega$ sur Ω. Il s'agit d'une extension du problème de Monge :

$$\inf\left\{\int|x-T(x)|^2d\mu\;:\;T:\Omega\to\Omega,\;T_\#\mu=\nu\right\}$$

(une extension parce qu'à toute fonction T de ce type on peut associer un plan γ_T en prenant $\gamma_T=(id\times T)_\#\mu$, tel que le coût de T dans le problème de Monge est le même que celui de γ_T dans celui de Kantorovitch ; de plus, sous des hypothèses opportunes sur μ, le minimum parmi les plans est atteint en un plan de la forme γ_T et, sous des hypothèses moins restrictives, l'ensemble des γ_T est dense dans $\Pi(\mu,\nu)$).

Indépendamment du fait que le minimum soit réalisé par une fonction T ou pas, on peut définir une distance sur les mesures, notée W_2, à l'aide de la valeur de ce minimum (7). Nous pouvons définir en effet la distance $W_2(\mu,\nu)$ comme la racine carrée

5. Des versions plus allégées de ces références existent, comme par exemple [AS], ou le tout récent *User's Guide to Optimal Transport* ([AG]), qui est en effet une très bonne référence pour cet exposé, puisqu'il traite à moitié de transport optimal (bien que le titre ne fasse référence qu'à cet aspect), puis pour un sixième de la théorie générale des flots de gradient (comme dans notre section 1), et pour un tiers des espaces métriques avec courbures bornées inférieurement (ce dont on parlera rapidement dans la section 3).

6. Cette terminologie est très contestée, en particulier en Russie, parce que ce L. Vaserstein, dont le nom est souvent écrit Wasserstein, n'a pas vraiment eu le rôle qu'on pourrait croire dans cette notion ; cependant, elle est tellement utilisée en Occident qu'il semble impossible de la changer au profit d'autres dénominations qui pourraient être plus appropriées, comme distance de Monge-Kantorovitch, ou autres...

7. L'indice 2 se réfère à l'exposant dans le coût quadratique, des distances W_p étant possibles pour $1\le p\le\infty$; la lettre W se réfère surtout au nom « Wasserstein », même si Kantorovitch utilisait déjà cette lettre, pour W = Work ; pour se débarrasser de cette terminologie controversée, et sans doute à la solde de l'industrie de l'entertainment, Y. Brenier proposait de passer plutôt à la notation MK_2, avec MK, bien évidemment, pour Monge-Kantorovitch...

de la valeur minimale :

$$W_2(\mu,\nu):=\left(\min\left\{\int_{\Omega\times\Omega}|x-y|^2\ d\gamma : \gamma\in\Pi(\mu,\nu)\right\}\right)^{1/2}=\min\big\{||x-y||_{L^2(\gamma)} : \gamma\in\Pi(\mu,\nu)\big\}.$$

On peut démontrer que W_2 est une distance sur $\mathcal{P}(\Omega)$, qui métrise la converge faible-$*$ des mesures de probabilité (même si des subtilités additionnelles se présentent dans le cas d'un domaine Ω non compact). On appellera donc *Espace de Wasserstein* $\mathcal{W}_2(\Omega)$ l'espace $\mathcal{P}(\Omega)$ muni de cette distance.

Pour une approche rapide et alternative aux flots de gradient dans l'espace de Wasserstein, nous présentons ici les ingrédients dont on a besoin à propos du problème du transport optimal et de W_2 pour les utiliser dans l'étude du schéma discret. Cette approche, basée sur des perturbations « verticales », différentes des perturbations « horizontales » utilisées dans [JKO] et [AGS05], a été essentiellement proposée par moi-même et ne représente que mon choix personnel ; j'ai commencé à l'utiliser dans d'autres contextes (voir [BS]) et l'ai ensuite appliquée aux flots de gradient (voir [MRS, San-X, San-O]).

Le point principal dont on aura besoin pour la suite est le suivant. Pour toute paire (μ,ν) il existe une fonction lipschitzienne $\varphi:\Omega\to\mathbb{R}$, appelée *potentiel de Kantorovitch* avec les propriétés suivantes :

– si φ est différentiable μ-p.p. (ce qui est le cas, par exemple, si $\mu\ll\mathcal{L}^d$), alors il y a une unique solution au problème de Kantorovitch, qui est de la forme γ_T, et le T optimal est donné par $T(x)=x-\nabla\varphi(x)$;
– la fonction φ joue aussi le rôle de dérivée (variation première) de $\frac{1}{2}W_2^2(\cdot,\nu)$: on a

$$\frac{d}{d\varepsilon}\frac{1}{2}W_2^2(\mu+\varepsilon\chi,\nu)_{|\varepsilon=0}=\int\varphi d\chi.$$

Cette fonction φ (définie à une constante additive près) est en fait la solution d'un problème d'optimisation « dual » du problème de Kantorovitch, au sens de l'analyse convexe (le problème de Kantorovitch étant un problème de programmation linéaire en dimension infinie, il admet une formulation duale, définie sur un espace de fonctions continues dont les mesures sont le dual).

Pour faciliter la suite de la lecture, on introduit la notation suivante pour les dérivées : étant donnée une fonctionnelle $G:\mathcal{P}(\Omega)\to\mathbb{R}$ on indique par $\frac{\delta G}{\delta\rho}(\rho)$, si elle existe, la seule fonction (à une constante additive près) telle que $\frac{d}{d\varepsilon}G(\rho+\varepsilon\chi)_{|\varepsilon=0}=\int\frac{\delta G}{\delta\rho}(\rho)d\chi$ pour toute perturbation χ telle que, au moins pour $\varepsilon\in[0,\varepsilon_0]$, la mesure $\rho+\varepsilon\chi$ soit dans $\mathcal{P}(\Omega)$. Ainsi, dans le cas du potentiel de Kantorovitch, la condition ci-dessus se lit $\frac{\delta\left(\frac{1}{2}W_2^2(\cdot,\nu)\right)}{\delta\rho}=\varphi$.

Un dernier ingrédient important concernant l'espace de Wasserstein est la connaissance de ses géodésiques, qui ont une forme assez intuitive. Étant donnés deux mesures $\mu,\nu\in\mathcal{P}(\Omega)$ et T un transport optimal de μ à ν, en supposant que le domaine Ω soit convexe, la courbe

$$\mu_t:=((1-t)id+tT)_\#\mu,\qquad t\in[0,1] \tag{7}$$

est une géodésique à vitesse constante reliant $\mu_0 = \mu$ à $\mu_1 = \nu$. De plus, toutes les géodésiques ont cette forme (ou une forme équivalente $\mu_t = (\pi_t)_\# \gamma$ où γ est un plan optimal et $\pi_t(x, y) = (1-t)x + ty$, au cas où l'optimum serait réalisé par un plan au lieu d'un transport). La convexité du domaine Ω sert à garantir $\mu_t \in \mathcal{P}(\Omega)$ et sera implicitement supposée dans le reste de la section.

2.2. Schéma discret dans l'espace de Wasserstein et EDP d'évolution

Nous pouvons maintenant considérer une fonctionnelle F définie sur l'espace de Wasserstein et définir le schéma discret

$$\rho_{k+1}^\tau \in \operatorname{argmin}_\rho F(\rho) + \frac{W_2^2(\rho, \rho^\tau(k))}{2\tau}.$$

Quelles sont les conditions d'optimalité ? De manière informelle, on peut écrire

$$\frac{\delta F}{\delta \rho}(\rho_{k+1}^\tau) + \frac{\varphi}{\tau} = \mathrm{const}$$

(la raison pour avoir une constante au lieu de 0, outre le fait que φ est aussi définie à une constante près, est que les conditions sur χ pour que $\rho + \varepsilon\chi \in \mathcal{P}(\Omega)$ imposent en particulier $\int d\chi = 0$). L'égalité ci-dessus peut en fait être démontrée proprement (voir [San-O]) et a lieu ρ-p.p. Nous allons passer directement aux conséquences de cette condition : si on met ensemble le fait que la somme ci-dessus est constante et que l'on a $T(x) = x - \nabla\varphi(x)$ pour le T optimal, on obtient

$$\frac{T(x) - x}{\tau} = -\frac{\nabla\varphi(x)}{\tau} = \nabla\Big(\frac{\delta F}{\delta \rho}(\rho)\Big)(x). \tag{8}$$

Nous allons écrire $-v$ pour le ratio $\frac{T(x)-x}{\tau}$. Pourquoi ? Parce que, en tant que ratio entre un déplacement et un pas de temps, il joue le rôle d'une vitesse, et le signe négatif est justifié par le fait qu'il représente le mouvement de ρ_{k+1}^τ à ρ_k^τ, en remontant donc le temps.

Or, comme on a obtenu $v = -\nabla\big(\frac{\delta F}{\delta\rho}(\rho)\big)$, cela suggère qu'à la limite $\tau \to 0$ on obtiendra une solution de l'équation

$$\frac{\partial \rho}{\partial t} - \nabla \cdot \Big(\rho \nabla\Big(\frac{\delta F}{\delta \rho}(\rho)\Big)\Big) = 0, \tag{9}$$

simplement parce que la densité ρ_t d'une famille de particules qui bouge instantanément à la vitesse v_t est solution (au sens des distributions) de l'*équation de continuité*

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0.$$

Cette équation est complémentée par des conditions au bord de Neumann [8] $v \cdot n_\Omega = 0$, correspondant au fait que la masse ne sort jamais du domaine Ω et permettant de vérifier la conservation de la masse.

8. Nous signalons juste qu'une approche proposée par Figalli et Gigli [FG] permet, en considérant une variante de la distance W_2 qui donne un rôle spécial au bord, d'obtenir des conditions de Dirichlet.

Pour clarifier le sens de (9), il est utile de donner des exemples. Nous considérons trois types de fonctionnelles sur $\mathcal{P}(\Omega)$:

$$F(\rho) = \int f(\rho(x))dx, \quad G(\rho) = \int V(x)d\rho, \quad H(\rho) = \frac{1}{2}\iint W(x-y)\rho(dx)\rho(dy),$$

où $f : \mathbb{R} \to \mathbb{R}$ est une fonction convexe et superlinéaire (et $\rho(x)$ indique la densité de ρ par rapport à la mesure de Lebesgue, la fonctionnelle F valant $+\infty$ si ρ n'est pas absolument continue), $V : \Omega \to \mathbb{R}$ et $W : \mathbb{R}^d \to \mathbb{R}$ sont des fonctions suffisamment régulières (on peut prendre W symétrique, par simplicité, $W(z) = W(-z)$). Dans ce cas il est assez simple de calculer

$$\frac{\delta F}{\delta \rho}(\rho) = f'(\rho), \quad \frac{\delta G}{\delta \rho}(\rho) = V, \quad \frac{\delta H}{\delta \rho}(\rho) = W * \rho.$$

Un exemple intéressant est celui de $f(t) = t \ln t$, qui donne $f'(t) = \ln t + 1$ et $\nabla(f'(\rho)) = \frac{\nabla \rho}{\rho}$: par conséquent, le flot de gradient associé à cette fonctionnelle F donnera l'*Équation de la Chaleur* $\frac{\partial \rho}{\partial t} - \Delta \rho = 0$. Si on prend $F + G$, on aura alors l'*Équation de Fokker-Planck* $\frac{\partial \rho}{\partial t} - \Delta \rho - \nabla \cdot (\rho \nabla V) = 0$. Le choix $f(t) = t^m$ peut donner l'*Équation des milieux poreux* et la fonctionnelle d'interaction H donne, au contraire, des équations non-locales (et non-linéares). Nous signalons aussi les équations pour les modèles de mouvement de foules, qui ont une forme différente et beaucoup moins régulière, étudiées d'abord dans un cas discret dans [MV1] (voir [MV2] pour plus de détails) et étendues au cas continu dans [MRS]. Ce n'est qu'en utilisant des méthodes de transport et flots de gradient que l'existence d'une solution dans le cas continu – auparavant trop dure – a été prouvée.

Sans rentrer dans les détails de comment démontrer que ce schéma discret converge à une solution de l'équation de continuité, nous donnons juste une petite idée supplémentaire, qui porte sur l'interpolation de la suite ρ_k^τ (et des vitesses discrètes correspondantes v_k^τ). En effet, deux interpolations différentes se révèlent utiles : d'une part, on peut définir une interpolation (ρ^τ, v^τ) constante par morceaux, comme dans (2) ; d'autre part, on peut connecter chaque mesure ρ_k^τ à ρ_{k+1}^τ en utilisant une géodésique $\tilde{\rho}^\tau$ de l'espace de Wasserstein, et (7) donne son expression explicite et montre qu'elle est obtenue en déplaçant les particules avec une vitesse $\tilde{v}^\tau$ apparentée à v^τ. Or, l'avantage de cette deuxième interpolation est qu'on a une courbe continue et qu'on peut garantir que l'équation de continuité est satisfaite. Au contraire, la première interpolation est discontinue et l'équation n'est pas satisfaite, mais il y a un lien entre v^τ et ρ^τ ($v^\tau = -\nabla\big(\frac{\delta F}{\delta \rho}(\rho^\tau)\big)$), ce qu'on n'a pas entre $\tilde{\rho}^\tau$ et $\tilde{v}^\tau$. Il est possible de démontrer que les deux interpolations convergent finalement à la même limite lorsque $\tau \to 0$, et que donc à la limite on satisfait l'équation de continuité avec un champ de vitesse qui est bien donné par $v = -\nabla\big(\frac{\delta F}{\delta \rho}(\rho)\big)$, ce qui permet de conclure.

Au niveau des applications en EDP, il me semble judicieux de souligner que

– le point important en EDP est que les courbes $(\rho_t)_t$ obtenues de cette manière sont de vraies solutions faibles de l'équation de continuité ; de ce point de vue,

le formalisme développé dans la première moitié de [AGS05] n'est pas crucial ; par contre, dans la deuxième moitié du livre les auteurs s'occupent exactement d'étudier les propriétés des courbes dans l'espace de Wasserstein (en donnant les résultats nécessaires pour pouvoir effectuer soigneusement le passage à la limite $\tau \to 0$ ci-dessus) et de faire l'équivalence entre les différentes notions. Pourtant, si le but est d'obtenir l'existence d'une solution faible, la théorie dans les espaces métriques est souvent excessivement lourde.
– Après avoir utilisé la théorie du transport optimal pour choisir la distance dans le schéma discret et ensuite pour l'interpolation, le passage à la limite peut se faire par des techniques classiques de compacité dans l'espace des mesures et d'analyse fonctionnelle.
– Les hypothèses de λ-convexité ne sont en général pas importantes à ce niveau, en ce qui concerne l'existence.
– Quant à l'unicité de la solution, le point de vue naturel serait celui de vérifier l'unicité des solutions faibles de l'équation (ou, éventuellement, de définir une notion plus restrictive de solution de l'EDP correspondante), ce qui est a priori différent des notions d'EDE ou d'EVI. Cependant, de fait l'unicité n'est pratiquement jamais prouvable sans des hypothèses de convexité.

2.3. Convexité géodésique pour W_2

Or, bien qu'on ait insisté sur le fait que dans ce cadre l'approche la plus naturelle est celle qui consiste à considérer les solutions faibles de certaines EDP, il est indéniable qu'il peut être intéressant de vérifier si la théorie développée à la section 1 peut s'appliquer aux fonctionnelles modèles F, G et H présentées ci-dessus, d'autant plus que cela peut être une manière de vérifier l'unicité. En particulier, on peut se demander si elles sont géodésiquement convexes, sachant qu'on connaît maintenant les géodésiques de l'espace de Wasserstein. Par exemple, il n'est pas difficile de vérifier que la convexité de V est suffisante pour garantir la convexité géodésique de G, puisque

$$G(\mu_t) = \int V d\big((1-t)id + tT\big)_{\#}\mu = \int V\big((1-t)x + tT(x)\big)d\mu,$$

tout comme la convexité de W garantit celle de H :

$$\begin{aligned}H(\mu_t) = \int W(x-y)\, d\left(\big((1-t)id + tT\big)_{\#}\mu\big) \otimes \big(\big((1-t)id + tT\big)_{\#}\mu\right)(x,y)\\ = \int W\big((1-t)x + tT(x), (1-t)y + tT(y)\big)d\mu\otimes\mu.\end{aligned}$$

De même, si V ou W sont λ-convexes on obtient une λ-convexité géodésique.

Aussi, on pourrait se demander si l'hypothèse $\mathrm{C}^2\mathrm{G}^2$ est satisfaite et, en regardant la question de près, on se rend compte que la définition a été justement donnée exprès pour faire face au cas de l'espace de Wasserstein. En effet, le problème principal de cet espace est que l'application $\mu \mapsto W_2^2(\mu,\nu)$ n'a pas de propriété de convexité géodésique mais elle est plutôt concave ; par contre, si on fixe $\nu \in \mathcal{P}(\Omega)$, et que l'on prend μ_0, μ_1 deux autres mesures de probabilité et T_0, T_1 les transports optimaux de ν vers μ_0 et μ_1, alors la courbe

$$\mu_t := ((1-t)T_0 + tT_1)_{\#}\nu \tag{10}$$

connecte μ_0 à μ_1 et peut être utilisée dans $\mathrm{C}^2\mathrm{G}^2$. Cette courbe est appelée une *géodésique généralisée* entre μ_0 et μ_1. Il n'est pas difficile de voir que $\nu \mapsto W_2^2(\mu_t, \nu)$ satisfait l'hypothèse de convexité que l'on cherche, puisque

$$W_2^2(\mu_t, \nu) \leq \int |x - ((1-t)T_0(x) + tT_1(x))|^2 \, d\nu$$

et le calcul suit de la convexité du coût quadratique. Aussi, la convexité des fonctionnelles G et H le long de la même courbe est évidente (puisqu'elle se basait uniquement sur le fait que la courbe μ_t était obtenue en interpolant linéairement deux fonctions de transport). La convexité de F est plus délicate, mais faisable, et tout est résumé dans la proposition suivante, essentiellement due à McCann ([McC]).

Proposition 2.1. — *Supposons que Ω est un domaine convexe de $\mathbb{R}^d$. Alors la fonctionnelle F est géodésiquement convexe dans l'espace de Wasserstein si f est convexe et $t \mapsto t^d f(t^{-d})$ est convexe décroissante. G et H sont λ-géodésiquement convexes si V et W sont λ-convexes, respectivement.*

Sous les mêmes hypothèses, la convexité de F, G et H est également vérifiée le long des géodésiques généralisées définies en (10).

Ce résultat permet d'appliquer à ces fonctionnelles la théorie de la section précédente.

Il est utile de remarquer que l'hypothèse sur la fonction f de la proposition 2.1 est satisfaite par toute fonction du type $f(t) = t^q$, $q > 1$, et par $f(t) = t \ln t$. Aussi, ce théorème se base, toujours en ce qui concerne la fonctionnelle F, sur la structure euclidienne de Ω et changerait profondément si on était sur une variété. Il a été signalé (voir [vRS]) que la condition de K-convexité géodésique de la fonctionnelle Entropie sur une variété caractérise une borne inférieure sur la courbure de la variété :

Proposition 2.2. — *Soient M une variété compacte de dimension d et ν sa mesure de volume renormalisée, de manière à avoir $\nu \in \mathcal{P}(M)$. Soit E la fonctionnelle entropie définie par $E(\rho) = \int \rho \ln \rho \, d\nu$ pour toute mesure $\rho \ll \nu$ (et $+\infty$ pour les mesures qui ne sont pas absolument continues par rapport à ν). Alors E est K-géodésiquement convexe dans l'espace de Wasserstein $(\mathcal{P}(M), W_2)$ si et seulement si la courbure de Ricci* Ric_M *satisfait* $\mathrm{Ric}_M \geq K$.

Pour le cas $K = 0$, la même équivalence est vraie si on remplace la fonction $f(t) = t \ln t$ par la fonction $f_N(t) = -t^{1-1/N}$ avec $N \geq d$.

Cela a donné lieu (on le verra dans la section 3) à une définition basée sur le transport optimal de la notion de courbure de Ricci dans des espaces plus abstraits, définition dont il est question dans les papiers célèbres de Sturm, [St], et, indépendamment, de Lott et Villani, [LV].

3. LE FLOT DE LA CHALEUR DANS LES ESPACES MÉTRIQUES

Cette troisième section se propose de ne présenter que le début des nouvelles recherches du trio Ambrosio-Gigli-Savaré, à nouveau centrées sur les espaces métriques, mais beaucoup plus ambitieuses, visant maintenant une théorie générale de l'analyse dans ces espaces, en tirant profit des techniques et des notions développées dans le cadre du transport optimal et des flots de gradient.

Un point de départ pour tout cela a été la théorie du flot de la chaleur qu'on tâchera de présenter brièvement ici. L'observation clé est la suivante : dans l'espace euclidien $\mathbb{R}^d$ (ou dans un domaine $\Omega \subset \mathbb{R}^d$), le flot de la chaleur $\partial_t \rho = \Delta\rho$ peut être vu comme un flot de gradient de deux manières différentes

– tout d'abord il s'agit du flot de gradient dans l'espace de Hilbert $L^2(\Omega)$, muni de la norme usuelle, de la fonctionnelle d'énergie de Dirichlet $\mathcal{D}(\rho) = \int |\nabla\rho|^2 dx$ (fonctionnelle qui vaut $+\infty$ si $\rho \notin H^1(\Omega)$) ; dans ce cadre, le point de départ ρ_0 pourrait être une fonction arbitraire dans $L^2(\Omega)$ mais les propriétés bien connues de l'équation de la chaleur nous disent que $\rho_0 \geq 0 \Rightarrow \rho_t \geq 0$ et, si Ω est tout l'espace, ou si les conditions au bord sont de Neumann, alors $\int \rho_0 dx = 1 \Rightarrow \int \rho_t dx = 1$, ce qui permet de se restreindre aux densités de probabilité.
– aussi, en prenant la fonctionnelle E de la section précédente (avec $f(t) = t \ln t$), le flot de la chaleur est également un flot de gradient dans l'espace $\mathcal{W}_2(\Omega)$, comme on l'a déjà remarqué.

La question naturelle est : est-ce que le fait que ces deux flots coïncident est un fait général ? Que faudrait-il faire pour analyser et donner une réponse à cette question dans un espace métrique général ? Dans l'espace euclidien c'est simple : on écrit l'EDP associée à ces deux flots et, comme c'est la même et qu'il y a unicité, les deux flots coïncident.

Tout d'abord on se rend compte que la question n'est pas bien posée si la seule structure dont on dispose est celle d'espace métrique, puisqu'il nous faut également une mesure de référence (rôle joué par la mesure le Lebesgue dans le cas euclidien), pour définir l'intégrale $\int |\nabla\rho|^2 dx$, ainsi que pour l'entropie $\int \rho \ln \rho \, dx$ (cette fonctionnelle n'étant pas 1-homogène, sa valeur dépend de la mesure par rapport à laquelle on calcule la densité ρ).

On décidera donc de se placer dans le cas des *espaces métriques mesurés*, (X, d, m), où $m \geq 0$ est une mesure (de préférence finie) sur les boréliens de X. Ceci ne doit pas surprendre, les espaces métriques mesurés sont en ce moment la nouvelle frontière de certaines branches de l'analyse géométrique, comme généralisation naturelle des variétés riemanniennes. Pour ne pas alourdir la liste des références, on peut se limiter à remarquer que les papiers suivants, déjà présents dans la bibliographie de cet exposé par ailleurs, touchent à ces questions : [AG, AGS-heat, AGS-riem, AGS-grad, Che, Gig10, GKO, Haj1, Haj2, HK, LV, Sha, St].

3.1. Énergies de Dirichlet et Cheeger dans des espaces métriques mesurés

Pour bien poser la question il faut d'abord définir le flot de l'Énergie de Dirichlet et donc définir cette énergie. En gros, cela revient à définir l'espace $H^1(X)$ pour X espace métrique mesuré (EMM). Ceci n'est pas une nouveauté, plusieurs auteurs l'ont étudié (en particulier [Haj1, Haj2, Che, Sha]) et les travaux récents d'Ambrosio, Gigli et Savaré ([AGS-heat, AGS-grad]) ont démontré des résultats complémentaires utiles à leur analyse en toute généralité (en particulier, beaucoup de résultats précédents demandaient des propriétés de *doubling* et l'existence d'une *inégalité de Poincaré* – voir aussi [HK] – comme hypothèse sur (X, d, m), absente dans ces papiers). L'une des premières définitions de l'espace de Sobolev sur un EMM avait été donnée par Haiłasz, qui utilisait la définition « $f \in H^1(X, d, m)$ si et seulement si il existe $g \in L^2(X, m)$ telle que $|f(x)-f(y)| \leq d(x, y)(g(x)+g(y))$ », propriété qui caractérise les espaces de Sobolev dans $\mathbb{R}^n$ en prenant pour g la fonction maximale de $|\nabla f|$. Cette définition n'étant pas locale, tous les travaux plus récents se fondent plutôt sur les idées de Cheeger ([Che]), basées sur la relaxation à partir des fonctions lipschitziennes, et de Shanmuganlingam ([Sha]), basées sur l'estimation $|f(x(0)) - f(x(1))| \leq \int |\nabla f(x(t)||x'(t)|$ demandée sur « presque toutes » courbes, en un sens opportun. Le papier récent [AGS-grad] présente une classification de différentes notions de « module du gradient » au sens faible dans un EMM et l'équivalence en analyse, mais ici on va se baser sur une seule définition de $\int |\nabla f|^2 dm$, celle qui nous semble la plus simple.

Pour toute fonction f lipschitzienne sur X, prenons sa *constante de Lipschitz locale* $|\nabla f|$ définie en (4), et posons $\mathcal{D}(f) := \int |\nabla f|^2(x) dm$. Ensuite, par relaxation on définit l'*Énergie de Cheeger*[9] $\mathcal{C}(f)$:

$$\mathcal{C}(f) := \inf \left\{ \liminf_n \ \mathcal{D}(f_n) \ : \ f_n \to f \text{ en } L^2(X, m), \ f_n \in \mathrm{Lip}(X) \right\}.$$

On peut ensuite définir l'espace de Sobolev $H^1(X, d, m)$ comme l'espace des fonctions telles que $\mathcal{C}(f) < +\infty$. Cet espace sera un espace de Banach muni de la norme $f \mapsto \sqrt{\mathcal{C}(f)}$ et la fonction C sera convexe. On peut définir $-\Delta f$ comme l'élément de norme minimale du sous-différentiel $\partial\mathcal{C}(f)$, sachant qu'en général l'application $f \mapsto -\Delta f$ ne sera pas linéaire (la norme $\sqrt{\mathcal{C}(f)}$ n'étant pas hilbertienne).

Définir le flot de la fonctionnelle C dans l'espace de Hilbert $L^2(X, m)$ n'est maintenant guère difficile, et rentre bien dans le cadre classique des fonctionnelles convexes dans les espaces de Hilbert et des opérateurs maximaux monotones (voir [Bre]), avec des résultats d'existence et d'unicité qui sont valables dans un cadre très général (et qui s'appliquent ici).

9. Le nom venant du fait que Cheeger avait aussi donné une définition par relaxation ; de plus, les auteurs ne souhaitaient pas l'appeler Énergie de Dirichlet, parce que ce nom indique habituellement une forme quadratique.

3.2. Un flot de gradient bien posé pour l'entropie

Une deuxième étape (développée d'abord dans [Gig10] et ensuite généralisée dans [AGS-heat]) consiste à donner des conditions d'existence et d'unicité du flot de l'entropie. Pour ce faire, on considère la fonctionnelle E, définie sur l'ensemble des densités f telles que $\rho := f \cdot m$ soit une mesure de probabilité, par $E(f) := \int f \ln f \, dm$ et on en considère les flots de gradient dans $\mathcal{W}_2$ au sens EDE. Pour pouvoir appliquer la théorie générale de la section 1, puisqu'on ne peut pas passer par la formulation faible de l'équation de continuité, il est naturel de supposer que cette fonctionnelle E soit λ-géodésiquement convexe pour un certain $\lambda \in \mathbb{R}$. Cela revient à dire, au sens de Sturm et Lott-Villani, que l'espace (X, d, m) est un EMM avec *courbure de Ricci minorée.* Nous donnons ici la définition correspondante, déjà évoquée en fin de section 2.

Définition 3.1. — *Un espace métrique mesuré (X, d, m) est dit avoir une courbure de Ricci minorée par $K \in \mathbb{R}$ au sens de Sturm et Lott-Villani si la fonctionnelle d'entropie $E : \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ définie par*

$$E(\rho) = \begin{cases} \int f \ln f dm & si\ \rho = f \cdot m \\ +\infty & si\ \rho\ n'est\ pas\ absolument\ continue\ par\ rapport\ à\ m \end{cases}$$

est K-géodésiquement convexe dans l'espace $\mathcal{W}_2(X)$. On dit alors que (X, d, m) satisfait la condition[10] *$CD(K, \infty)$.*

Remarquons aussi que la formulation EVI du flot de gradient n'est pas disponible dans ce cas, parce qu'il nous manque la convexité du carré de la distance de Wasserstein le long des géodésiques. Pour pouvoir bien définir le flot de gradient de E, Gigli a introduit dans [Gig10] une stratégie intéressante de preuve d'unicité pour les flots EDE.

Proposition 3.2. — *Si $F : \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ est une fonctionnelle strictement convexe (par rapport aux combinaisons convexes usuelles $\mu_s := (1-s)\mu_0 + s\mu_1$, ce qui a un sens dans le cadre des mesures de probabilités), telle que $|\nabla^- F|$ est un gradient supérieur pour F et que son carré $|\nabla^- F|^2$ est convexe, alors pour toute mesure $\bar{\mu}$ de départ il existe au plus un seul flot de gradient $\mu(t)$ au sens EDE pour la fonctionnelle F démarrant en $\bar{\mu}$: $\mu(0) = \bar{\mu}$.*

En particulier, cela s'applique au cas de l'entropie E (la stricte convexité de la fonctionnelle étant évidente, la convexité de la pente se basant sur la formulation (6), valable en cas de λ-convexité géodésique, et sur des propriétés algébriques particulières de la fonction $t \mapsto t \ln t$).

10. On utilise la notation générale $CD(K, N)$, qui représente le fait d'avoir une courbure minorée par K et une dimension majorée par N (« CD » signifiant « curvature-dimension »).

3.3. Comparaison des flots de gradient

Il reste ensuite à vérifier (et ce n'est pas trivial), que tout flot de gradient de $\mathcal{C}$ (par rapport à la métrique L^2) est aussi un flot de gradient EDE pour E et la métrique W_2. Celle-ci $((\mathcal{C}, L^2) \Rightarrow (E, W_2))$ est la direction la plus simple dans ce cadre, puisque les calculs sont plus simples à faire, la structure des flots de gradient des fonctionnelles convexes dans les espaces de Hilbert étant plus claire. Pour ce faire, il convient de calculer et d'estimer

$$\frac{d}{dt}E(f_t), \quad \text{où } f_t \text{ est un flot de gradient de } \mathcal{C} \text{ dans } L^2(X, m).$$

Ceci se base essentiellement sur une stratégie développée dans [GKO] et sur un lemme dû à Kuwada. La preuve initiale, contenue dans [GKO], est valable dans le cas des espaces d'Alexandroff [11]. La généralisation de ce même résultat aux EMM arbitraires satisfaisant $CD(K, \infty)$ est faite dans [AGS-heat].

Proposition 3.3. — *Si f_t est un flot de gradient de $\mathcal{C}$ dans $L^2(X, \mathcal{H}^n)$, on a l'égalité avec l'information de Fischer :*

$$-\frac{d}{dt}E(f_t) = \mathcal{C}(2\sqrt{f_t}).$$

De plus, pour tout $\rho = f \cdot \mathcal{H}^n \in \mathcal{P}(X)$ on a

$$\mathcal{C}(2\sqrt{f}) \geq |\nabla^- E|^2(\rho)$$

(où la pente de E est calculée par rapport à la distance W_2). Aussi, en prenant la courbe $\rho_t = f_t \cdot \mathcal{H}^n$, on trouve que ρ_t est une courbe absolument continue dans l'espace $\mathcal{W}_2(X)$ et

$$|\rho'|(t)^2 \leq \mathcal{C}(2\sqrt{f_t}).$$

Ces trois estimations impliquent que ρ_t est un flot de gradient pour l'entropie E par rapport à W_2.

Une fois qu'on a cette équivalence, on peut se poser la question des propriétés de ce flot de gradient. La distance L^2 étant hilbertienne, il est facile de vérifier la propriété $\mathrm{C}^2\mathrm{G}^2$ qui permette de dire que ce flot satisfait la condition EVI. Au contraire, il n'est pas du tout évident qu'il la satisfasse en ce qui concerne la fonctionnelle E et la distance W_2. On peut vérifier que les trois assertions suivantes (dont le caractère vrai ou faux dépend de l'espace (X, d, m), qui est supposé satisfaire $CD(K, \infty)$) sont équivalentes (voir [AGS-riem]) :

11. Il s'agit (voir [BGP]) d'espaces métriques où les triangles sont au moins aussi « gros » (en comparant les distances d'un sommet aux points d'une géodésique connectant les autres sommets) que les triangles d'une variété modèle de comparaison de courbure constante k en dimension 2 (la dimension 2 étant suffisante, puisque seuls les triangles apparaissent dans la définition). Il se trouve que ces espaces ont toujours une dimension entière, et on peut les considérer comme des EMM quand ils sont de dimension finie, et on les munit de la mesure $\mathcal{H}^n$.

– l'unique flot de gradient de l'entropie E dans l'espace de Wasserstein satisfait la condition EVI ;
– le flot de la chaleur (qui est en même temps flot de E pour W_2 et de $\mathcal{C}$ pour L^2) est linéaire par rapport aux données initiales ;
– (si on suppose de plus que (X, d, m) est une variété finslérienne munie de sa distance naturelle et de sa mesure de volume) X est une variété riemannienne.

Par conséquent, Ambrosio, Gigli et Savaré ont proposé dans [AGS-riem] une définition d'EMM qui a une *courbure de Ricci riemannienne* bornée inférieurement, en demandant à satisfaire $CD(K, \infty)$ et la linéarité du flot de la chaleur. C'est la notion d'espace « infinitésimalement hilbertien » mentionnée dans la section 1.

Il est important de remarquer (mais nous ne développons pas ce sujet ici) que ces notions de bornes sur la courbure (riemanniennes ou pas) sont stables par convergence de Gromov-Hausdorff mesurées (une convergence à la Gromov-Hausdorff basée sur la distance de Wasserstein minimale entre les mesures images des deux espaces par des plongements isométriques dans un même espace).

RÉFÉRENCES

[Amb] L. AMBROSIO – *Movimenti minimizzanti*, Rend. Accad. Naz. Sci. XL Mem. Mat. Sci. Fis. Natur. **113** (1995), 191–246.

[AG] L. AMBROSIO et N. GIGLI – *A user's guide to optimal transport*, prépublication, 2011.

[AGS05] L. AMBROSIO, N. GIGLI et G. SAVARÉ – *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Math., ETH Zürich, 2005.

[AGS-heat] L. AMBROSIO, N. GIGLI et G. SAVARÉ – *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, prépublication, 2011.

[AGS-riem] L. AMBROSIO, N. GIGLI et G. SAVARÉ – *Metric measure spaces with Riemannian Ricci curvature bounded from below*, prépublication, 2011.

[AGS-grad] L. AMBROSIO, N. GIGLI et G. SAVARÉ – Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces, à paraître dans Rev. Mat. Iberoamericana, 2013.

[AGS-comp] L. AMBROSIO, N. GIGLI et G. SAVARÉ – *Heat flow and calculus on metric measure spaces with Ricci curvature bounded below - the compact case*, prépublication, 2012.

[AS] L. AMBROSIO et G. SAVARÉ, *Gradient flows of probability measures*, Handbook of differential equations, Evolutionary equations **3**, ed. by C.M. Dafermos and E. Feireisl, Elsevier, 2007.

[Bre] H. BREZIS – *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland mathematics studies, 1973.

[BGP] Yu. BURAGO, M. GROMOV, et G. PERELMAN – *A. D. Alexandrov spaces with curvatures bounded below* (Russian), Uspekhi Mat. Nauk **47** (1992), 3–51 ; English translation : Russian Math. Surveys **47** (1992), 1–58.

[BS] G. BUTTAZZO et F. SANTAMBROGIO – A model for the optimal planning of an urban area, *SIAM J. Math. Anal.* **37** (2005), 514–530.

[Che] J. CHEEGER – *Differentiability of Lipschitz functions on metric measure spaces,* Geom. Funct. Anal. **9** (1999), 428–517.

[DS] S. DANERI et G. SAVARÉ – *Eulerian calculus for the displacement convexity in the Wasserstein distance*. SIAM J. Math. Analysis **40** (2008), 1104-1122.

[DeG] E. DE GIORGI – *New problems on minimizing movements*, Boundary Value Problems for PDE and Applications, C. Baiocchi and J.-L. Lions, Masson, 1993, 81–98.

[FG] A. FIGALLI et N. GIGLI – *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions* J. Math. Pures Appl. **94** (2010), 107–130.

[Gig10] N. GIGLI – *On the Heat flow on metric measure spaces : existence, uniqueness and stability*, Calc. Var. Part. Diff. Eq. **39** (2010), 101–120.

[Gig-HDR] N. GIGLI – *Propriétés géométriques et analytiques de certaines structures non lisses*, Mémoire HDR, Univ. Nice-Sophia-Antipolis, 2011.

[GKO] N. GIGLI, K. KUWADA et S. I. OHTA – *Heat flow on Alexandrov spaces*, Comm. Pure Appl. Math., publié en ligne, sous presse, 2012.

[Haj1] P. HAJŁASZ – *Sobolev spaces on an arbitrary metric space*, Potential Analysis **5** (1996), 403–415.

[Haj2] P. HAJŁASZ – *Sobolev spaces on metric-measure spaces*, Contemp. Math. **338** (2003), 173–218.

[HK] P. HAJŁASZ et P. KOSKELA – *Sobolev met Poincaré*, Mem. Amer. Math. Soc. **688** (2000), 1–101.

[JKO] R. JORDAN, D. KINDERLEHRER et F. OTTO – *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), 1–17.

[Kan] L. V. KANTOROVITCH – *On the transfer of masses*, Dokl. Akad. Nauk. SSSR **37** (1942), 227–229.

[LV] J. LOTT et C. VILLANI – *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. **169** (2009), 903–991.

[MRS] B. MAURY, A. ROUDNEFF-CHUPIN et F. SANTAMBROGIO – *A macroscopic crowd motion model of gradient flow type*, Math. Mod. Meth. Appl. Sci. **20** (2010), 1787–1821.

[MV1] B. MAURY et J. VENEL – *A mathematical framework for a crowd motion model*, C. R. Acad. Sci. Paris, Ser. I **346** (2008), 1245–1250.

[MV2] B. MAURY et J. VENEL – *A discrete contact model for crowd motion*, ESAIM : M2AN **45** (2011), 145-168.

[vRS] M.-K. von RENESSE et K.-T. STURM – *Entropic measure and Wasserstein diffusion*, Ann. Probab. **37** (2009), 1114-1191.

[McC] R. J. MC CANN – *A convexity principle for interacting gases*, Adv. Math. **128** (1997), 153–159.

[Otto] F. OTTO – *The geometry of dissipative evolution equations : the porous medium equation*, Comm. Partial Differential Equations **26** (2001), 101–174.

[San-X] F. SANTAMBROGIO – *Gradient flows in Wasserstein spaces and applications to crowd movement*, Séminaire Laurent Schwartz no 27, École Polytechnique, 2010.

[San-O] F. SANTAMBROGIO – *Optimal transport for applied mathematicians*, notes de cours d'École doctorale, Orsay, 2012, en révision, disponibles à la page `http://www.math.u-psud.fr/∼santambr/OTAM.pdf`

[Sha] N. SHANMUGALINGAM – *Newtonian spaces : an extension of Sobolev spaces to metric measure spaces*, Rev. Mat. Iberoamericana **16** (2000), 243–279.

[St] K.-T. STURM – *On the geometry of metric measure spaces. I–II*, Acta Math. **196** (2006), 65–177.

[Vil03] C. VILLANI – *Topics in optimal transportation*, Grad. Stud. Math. **58**, AMS, Providence, 2003.

[Vil09] C. VILLANI – *Optimal transport, old and new*, Grund. Math. Wiss. **338**, 2009.

Filippo SANTAMBROGIO

Université Paris-Sud
Laboratoire de Mathématiques d'Orsay
UMR 8628 du CNRS
Bâtiment 425
91405 Orsay Cedex
E-mail : `filippo.santambrogio@math.u-psud.fr`