Séminaire BOURBAKI Janvier 2013
65e année, 2012-2013, n° 1065, p. 219 à 243

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

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

l'inégalité venant de la propriété de monotonie $(\nabla F(x) - \nabla F(y)) \cdot (x - y) \geqslant 0$, 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 $(\nabla F(x) - \nabla F(y)) \cdot (x - y) \geqslant \lambda |x - y|^2$, et donc

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

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 $t = 0, \tau, 2\tau, \ldots, k\tau, \ldots$. 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_{k+1}^τ nous donne

$$F(x_{k+1}^\tau) + \frac{d(x_{k+1}^\tau, x_k^\tau)^2}{2\tau} \leqslant F(x_k^\tau), \tag{3}$$

ce qui entraîne

$$d(x_{k+1}^\tau, x_k^\tau)^2 \leqslant 2\tau \left(F(x_k^\tau) - F(x_{k+1}^\tau)\right).$$

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

$$\sum_{k=0}^{l} d(x_{k+1}^\tau, x_k^\tau)^2 \leqslant 2\tau \left(F(x_0^\tau) - F(x_{l+1}^\tau)\right) \leqslant C\tau.$$

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 $|l - k| \leqslant \frac{|t-s|}{\tau} + 1$)

$$\begin{aligned} d(x^\tau(t), x^\tau(s)) \leqslant \sum_{k=0}^{l} d(x_{k+1}^\tau, x_k^\tau) \leqslant \left(\sum_{k=0}^{l} d(x_{k+1}^\tau, x_k^\tau)^2\right)^{1/2} \left(\frac{|t-s|}{\tau} + 1\right)^{1/2} \\ \leqslant C\big(|t-s|^{1/2} + \sqrt{\tau}\big). \end{aligned}$$

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

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

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

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

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

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

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

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

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.