Transport optimal sur une variété Riemannienne et applications

N. Gozlan

11 novembre 2011

1 Introduction

Le but de cet exposé est de donner une preuve du théorème de McCann caractérisant le transport optimal quadratique sur une variété Riemannienne.

Théorème 1.1 (McCann (2001))**.** *Soit μ une probabilité absolument continue par rapport à la mesure de volume sur une variété Riemannienne compacte M. Pour toute probabilité ν sur M, il existe une application T définie μ presque partout par*

$$T(x) = \exp_x(-\nabla\psi(x))$$

où ψ est une fonction différentiable presque partout, telle que $T\#\mu = \nu$ et

$$\int d^2(x, T(x))\,\mu(dx) = \inf_{St.qS\#\mu=\nu} \int d^2(x, S(x))\,\mu(dx).$$

Nous énoncerons plus loin une version plus précise de ce théorème qui généralise celui de Brenier à des espaces courbés. Nous verrons ensuite comment ce théorème peut être utiliser pour généraliser l'inégalité de Prekopa-Leindler à des variétés (travaux de Cordero-Erausquin - McCann - Schmuckenschläger).

Les résultats mentionnés ci-dessus sont issus des articles suivants :

• McCann, R. *Polar factorization of maps on Riemannian manifolds.* Geom. Funct. Anal. 11 (2001), no. 3, 589–608.

• Cordero-Erausquin D., McCann R., Schmuckenschläger M. *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math. 146 (2001), 219–257.

• Cordero-Erausquin D., McCann R., Schmuckenschläger M. *Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields and optimal transport*, Ann. Fac. Sci. Toulouse Math. (6) 15 (2006), no. 4, 613-635.

On recommande par ailleurs la lecture de

• Cordero-Erausquin D., *Quelques exemples d'applications du transport de mesure en géométrie euclidienne et riemannienne*, Séminaire de théorie spectrale et géométrie, Grenoble, volume 22 (2004), 125–152.

• Figalli A. and Villani C., *Optimal Transport and Curvature*, Nonlinear Pde's and Applications. Lectures from the C.I.M.E. Summer School held in Cetraro, June 23-28, 2008. Lecture Notes in Mathematics 2028, 2011.

2 Quelques notions sur les variétés Riemanniennes

Nous nous bornerons à poser quelques définitions et à rappeler sans preuve quelques théorèmes utiles pour la suite. Toutes ces notions pourraient faire l'objet d'un exposé plus fouillé lors d'une prochaine séance.

2.1 Variétés différentiables, espace tangents, champs de vecteurs, etc...

Définition 2.1.
1. *Une* variété topologique M de dimension *n est un espace topologique séparé tel que pour tout $p \in M$, il existe un ouvert U contenant p, un ouvert U' de $\mathbb{R}^n$ et un homéomorphisme $\varphi : U \to U'$. On dit que (U, φ) est une* carte, *U est le domaine de la carte.*
2. *Un* atlas *sur M est une famille de cartes $\mathcal{A}$, tel que pour tout p soit dans le domaine d'au moins une carte de $\mathcal{A}$.*
3. *Un* atlas *est dit de classe $\mathcal{C}^k$ si pour toutes cartes (U_1, φ_1), (U_2, φ_2) dont les domaines sont d'intersection non vide, le changement de carte $\varphi_2 \circ \varphi_1^{-1}$ est un difféomorphisme de classe $\mathcal{C}^k$.*
4. *Une variété de classe $\mathcal{C}^k$ est un couple $(M, \mathcal{A})$ où $\mathcal{A}$ est un atlas de classe $\mathcal{C}^k$ sur M.*

Dans la suite, on supposera toujours que M est une variété de classe $\mathcal{C}^\infty$.

Définition 2.2 (Applications différentiables)**.** *Une application $f : M \to N$ entre deux variétés différentiables est dite de classe $\mathcal{C}^k$ si pour toutes cartes (U, φ) de M et (V, ψ) de N, l'application $\psi \circ f \circ \varphi^{-1}$ est de classe $\mathcal{C}^k$ (là où elle est bien définie).*

Définition 2.3 (Espace tangent)**.** *Soit $p \in M$; une dérivation au point p est une application linéaire $\delta : \mathcal{C}^\infty(M) \to \mathbb{R}$ telle que $\delta(fg) = f(p)\delta(g) + g(p)\delta(f)$. L'ensemble noté T_pM de toutes les dérivations au point p est appelé* espace tangent *en p.*

Proposition 2.4. *Pour tout $p \in M$, T_pM est un espace vectoriel de dimension n.*

Exemple : Champs de base. Soit (U, φ) une carte. Pour tout $p \in U$, on définit $X_p^i \in T_pM$ par

$$X_p^i[f] = \frac{\partial f \circ \varphi^{-1}}{\partial x_i}(\varphi(p)), \qquad \forall f \in \mathcal{C}^\infty(M).$$

Les vecteurs $X_p^1, X_p^2, \dots, X_p^n$ forment une base de T_pM.

Exemples : Vecteurs tangents d'une courbe. Soit $\alpha : I \to M$ un chemin de classe $\mathcal{C}^1$. Pour tout $t \in I$, on définit le vecteur $\dot{\alpha}(t) \in T_{\alpha(t)}M$ par la formule

$$\dot{\alpha}(t)[f] = \frac{d}{dt} f \circ \alpha(t), \qquad \forall f \in \mathcal{C}^\infty(M).$$

Définition 2.5 (Application tangente). *Si $f : M \to N$ est une fonction lisse, pour tout $p \in M$, la différentielle de f en p, notée $d_p f : T_pM \to T_{f(p)}N$ définie par*

$$d_p f(v)[h] = v[h \circ f], \qquad \forall v \in T_pM, \qquad \forall h \in \mathcal{C}^\infty(M).$$

Définition 2.6 (Champ de vecteurs). *Soit D un ouvert de M ; un champ de vecteurs Y sur D est une application qui à un point $p \in D$ associe un vecteur $Y_p \in T_pM$.*

Exemple : Les vecteurs $X^1, \dots, X^n$ associés à une carte (U, φ) sont des champs de vecteurs sur U. De plus, si Y est un champ de vecteur sur U, alors il s'écrit d'une unique manière sous la forme

$$Y = \sum_{i=1}^{n} a_i X^i,$$

où les a_i sont des fonctions définies sur U.

Définition 2.7 (Champ de vecteurs le long d'une courbe). *Soit $\alpha : I \to M$ un chemin ; un champ de vecteur le long de α est une application Y qui à $t \in I$ associe $Y(t) \in T_{\alpha(t)}M$.*

Exemple : Si α est un chemin de classe $\mathcal{C}^1$, $t \mapsto \dot{\alpha}(t)$ est un champ de vecteurs le long de α.

Ecriture en coordonnées locales : Si (U, φ) est une carte et α est à valeurs dans U, un champ de vecteurs Y le long de α s'écrit d'une unique manière sous la forme

$$Y(t) = \sum_{i=1}^{n} a_i(t) X^i_{\alpha(t)}, \qquad \forall t \in I.$$

2.2 Variétés Riemanniennes

2.2.1 Métrique Riemannienne

Définition 2.8 (Variété Riemannienne). *Une métrique Riemannienne g sur une variété différentiable M est une application qui à tout point $p \in M$ associe un produit scalaire $g_p(\cdot, \cdot)$ sur T_pM. On impose que si Y, Z sont deux champs de vecteurs lisses, alors $p \mapsto g_p(Y, Z)$ est une fonction lisse. Un couple (M, g) est appelé* variété Riemannienne.

Soit (U, φ) une carte et X^i les champs de bases associés ; la *première forme fondamentale* sont les fonctions $g_{i,j}$ définies sur U par

$$g_{i,j}(p) = \langle X^i_p, X^j_p \rangle_p, \qquad \forall p \in U.$$

Exemple [vecteur gradient] : Si $f : M \to \mathbb{R}$ est lisse, alors pour tout $p \in M$, $d_p f$ est une forme linéaire sur T_pM. Donc il existe un unique vecteur noté $\nabla f(p)$ tel que

$$d_p f(v) = \langle \nabla f(p), v \rangle_p, \qquad \forall v \in T_pM.$$

Théorème 2.9. *Pour tous $x, y \in M$, on pose*

$$d(x,y) = \inf \left\{ \int_0^1 |\dot{\alpha}(t)| \, dt ; \alpha(0) = x,\ \alpha(1) = y,\ \alpha \text{ de classe } \mathcal{C}^1 \right\}.$$

La fonction d est une distance sur M qui redonne la topologie de M.

2.2.2 Forme volume

On définit une mesure positive sur M notée vol et appelée forme volume sur M de la manière suivante. On prend (U,φ) une carte, et on pose pour toute fonction f continue à support compact dans U

$$I(f,\varphi)=\int_{\varphi(U)} f\circ\varphi^{-1}(x)\sqrt{\det(G)}(\varphi^{-1}(x))\,dx_1\cdots dx_n, \qquad \text{où } G=[g_{i,j}]_{i,j}.$$

On constate, en utilisant la formule du changement de variable dans $\mathbb{R}^n$, que si (U,ψ) est une autre carte, alors $I(f,\varphi)=I(f,\psi)=I$. On pose $\int_U f\,d\mathrm{vol}=I$. On obtient une mesure sur M tout entier en utilisant une partition de l'unité.

2.2.3 Dérivée covariante.

On peut définir un procédé de dérivation des champs de vecteurs le long de courbes, appelé dérivation covariante de Levi - Civita.

Soit $\alpha: I\to M$ un chemin régulier et Y un champ de vecteurs régulier le long de α. La dérivation covariante de Y le long de α est notée $\frac{D}{dt}Y$ et est à nouveau un champ de vecteur le long de α :

$$\forall t\in I, \qquad \frac{D}{dt}Y(t)\in T_{\alpha(t)}M.$$

Cette dérivation vérifie les règles de calculs suivantes :
Si Y,Z sont deux champs de vecteurs le long d'une courbe $\alpha: I\to\mathbb{R}$, on a

- $\frac{D}{dt}(Y+Z)=\frac{D}{dt}Y+\frac{D}{dt}Z$.
- $\frac{D}{dt}(aY)=a'(t)Y+a\frac{D}{dt}Y$, où $a: I\to\mathbb{R}$ est une fonction régulière.
- $\frac{d}{dt}\langle Y,Z\rangle=\langle\frac{D}{dt}Y,Z\rangle+\langle Y,\frac{D}{dt}Z\rangle$.
- Si $\sigma: I\times J\to M:(s,t)\mapsto\sigma(s,t)$ est une fonction régulière, alors
$$\frac{D}{\partial t}\frac{\partial\sigma}{\partial s}(s,t)=\frac{D}{\partial s}\frac{\partial\sigma}{\partial t}(s,t).$$

Si M est une sous variété de dimension n de l'espace $\mathbb{R}^N$, cette dérivation peut se définir très simplement par la formule :

$$\frac{D}{dt}Y=\Pi_{\alpha(t)}\left(\frac{d}{dt}Y\right),$$

où pour tout $p\in M$, Π_p désigne la projection orthogonale sur l'espace tangent au point p, et où $\frac{d}{dt}Y$ est la dérivée au sens usuel de la fonction vectorielle à valeurs dans $\mathbb{R}^N$ $t\mapsto Y(t)$.

Exemple : Si α est un chemin régulier, son accélération est par définition $\frac{D}{dt}\dot\alpha(t)$.

Expression en coordonnées locales. Soit (U,φ) une carte, α un chemin régulier à valeurs dans U et $Y(t)=\sum_{i=1}^n b_i(t)X^i_{\alpha(t)}$ un champ de vecteurs le long de α. En notant $\varphi(\alpha(t))=(a_1(t),\ldots,a_n(t))\in\mathbb{R}^n$, on a la formule

$$\frac{D}{dt}Y(t)=\sum_{k=1}^n\left[\dot{b_k}(t)+\sum_{i,j}\Gamma^k_{i,j}(\alpha(t))b_j(t)\dot{a_i}(t)\right]X^k_{\alpha(t)}.$$

Les nombres Γ_{ij}^k sont les *symboles de Christoffel* définis par les formules

$$\Gamma_{ij}^k(p) = \sum_{l=1}^n \Gamma_{ijl}(p) g^{lk}(p),$$

où la matrice $[g^{kl}]$ est l'inverse de la matrice $[g_{ij}]$ et où

$$\Gamma_{ijl}(p) = \frac{1}{2}\left(\frac{\partial g_{jl}\circ\varphi^{-1}}{\partial x_i} + \frac{\partial g_{li}\circ\varphi^{-1}}{\partial x_j} - \frac{\partial g_{ij}\circ\varphi^{-1}}{\partial x_l}\right)(\varphi(p))$$

2.2.4 Géodésiques.

Définition 2.10. *Une géodésique est une courbe $\gamma : I \to M$ telle que $\frac{D}{dt}\dot\gamma(t) = 0$ pour tout $t \in I$.*

Les géodésiques sont automatiquement paramétrées à vitesse constante :

$$\frac{d}{dt}\langle\dot\gamma,\dot\gamma\rangle = 2\left\langle \frac{D}{dt}\dot\gamma,\dot\gamma\right\rangle = 0.$$

Equation en coordonnées des géodésiques : Soit (U,φ) une carte ; la courbe $\gamma : I \to M$ est une géodésique si et seulement si ces coordonnées $(\gamma_1,\dots,\gamma_n) = \varphi(\gamma) \in \mathbb{R}^n$ vérifient le système d'équations différentielles suivant :

$$\forall k \in \{1,\dots,n\}, \qquad \gamma_k''(t) + \sum_{i,j}\Gamma_{i,j}^k(\gamma(t))\gamma_i'(t)\gamma_j'(t) = 0$$

Théorème 2.11. *Pour tout $p \in M$ et $v \in T_pM$, il existe une unique géodésique maximale telle que $\gamma(0) = p$ et $\dot\gamma(0) = v$. L'espace métrique (M,d) est complet si et seulement si les géodésiques maximales sont définies sur tout $\mathbb{R}$.*

Théorème 2.12 (Hopf-Rinow). *Les propositions suivantes sont équivalentes*

1. *L'espace métrique (M,d) est complet.*
2. *Les géodésiques maximales sont définies sur $\mathbb{R}$ tout entier.*
3. *Les parties fermées bornées de M sont compactes.*

Dans ces conditions, pour tous $x,y \in M$, il existe au moins une géodésique $\gamma : [0,1] \to \mathbb{R}$ telle que $\gamma(0) = x$ et $\gamma(1) = y$ et réalisant l'infimum dans $d(x,y)$. On appelle ces courbes géodésiques minimisantes.

Application exponentielle. On suppose que la variété est complète. Soit $p \in M$, $v \in T_pM$; on définit $\exp_p(v) = \gamma(1)$, où γ est l'unique géodésique maximale telle que $\gamma(0) = p$ et $\dot\gamma(0) = v$. On a alors pour tout $t \in \mathbb{R}$, $\exp_p(tv) = \gamma(t)$. En effet, la courbe $\sigma(s) = \gamma(st)$, $s \in \mathbb{R}$ vérifie $\dot\sigma(0) = tv$ et vérifie l'équation des géodésiques. Donc $\sigma(1) = \exp_p(tv)$.

3 Le théorème de Mc Cann

On suppose que M est une variété Riemannienne connexe compacte. La distance Riemannienne d est donnée par

$$d(x,y) = \inf_{\gamma(0)=x,\gamma(1)=y}\int_0^1 |\dot\gamma(t)|\,dt,$$

où l'infimum porte sur les courbes de classe $\mathcal{C}^1$ joignant x à y.
Dans la suite, nous noterons

$$c(x,y) = \frac{1}{2}d^2(x,y).$$

Le coût de transport quadratique entre deux mesures de probabilités μ et ν sur M est

$$\mathcal{T}_2(\mu,\nu) = \inf_{\pi\in P(\mu,\nu)} \iint c(x,y)\,\pi(dxdy),$$

où l'ensemble $P(\mu,\nu)$ est défini par

$$P(\mu,\nu) = \{\pi \in \mathcal{P}(M^2); \pi(dx\times M) = \mu(dx) \text{ et } \pi(M\times dy) = \nu(dy)\}.$$

On sait, par un argument simple de compacité, que cet infimum est atteint. Le but est de trouver la forme des couplages optimaux.

D'après le théorème de dualité de Kantorovich vu lors du dernier exposé, on a l'identité suivante

$$\mathcal{T}_2(\mu,\nu) = \sup_{(\psi,\varphi)\in\Phi_c} \int \psi\,d\mu + \int \varphi\,d\nu.$$

L'ensemble Φ_c est constitué de tous les couples (ψ,φ) de fonctions continues sur M telles que

$$\psi(x) + \varphi(y) \le c(x,y), \qquad \forall x,y \in M.$$

Dans un premier temps, nous allons montrer que le problème dual possède une solution.

Définition 3.1 (c-conjugaison)**.** *Si $f : M \to \mathbb{R}$ est une fonction bornée, on pose*

$$f^c(y) = \inf_{x\in M}\{c(x,y) - f(x)\}.$$

On dit que f est c-concave si elle vérifie $f^{cc} = f$.

On voit facilement qu'on a toujours $f^{ccc} = f^c$. Donc f^c est toujours c-concave.

Proposition 3.2. *On note $L = \sup_{x,y\in M} d(x,y)$, et $J_c(u,v) = \int u\,d\mu + \int v\,d\nu$.*

1. *La fonction f^c est L-Lipschitz.*
2. *Si $(u,v)\in\Phi_c$ alors $(v^c,v^{cc})\in\Phi_c$, et $J_c(u,v)\le J_c(v^c,v^{cc})$.*

Démonstration. (1) Comme un inf de fonctions L-Lipschitz est encore L-Lipschitz, il suffit de montrer que, pour tout $y\in M$, la fonction $x\mapsto c(x,y)$ est L-Lispchitz. Pour tous $x,y,z\in M$, on a

$$\begin{aligned}\frac{1}{2}d^2(x,y) - \frac{1}{2}d^2(z,y) &= \frac{1}{2}\left(d(x,y)-d(z,y)\right)\left(d(x,y)+d(z,y)\right)\\ &\le L d(x,z),\end{aligned}$$

en utilisant l'inégalité triangulaire et la définition de L.

(2) Par définition, $v(y) + v^c(x) \le c(x,y)$ donc le couple $(v^c,v)\in\Phi(c)$. Par le même argument on conclut que $(v^c,v^{cc})\in\Phi_c$. Par ailleurs, l'inégalité $u(x)+v(y)\le c(x,y), \forall x,y$ entraîne que $u\le v^c$ et donc que $J_c(u,v)\le J_c(v^c,v)$. Comme $v\le v^{cc}$, on a aussi $J_c(v^c,v)\le J_c(v^c,v^{cc})$. □

Théorème 3.3. *Pour toutes probabilités μ, ν sur M, il existe une fonction ψ c-concave telle que*

$$\sup_{(u,v)\in\Phi_c} J_c(u,v) = J_c(\psi,\psi^c).$$

Démonstration. Soit $(u_n, v_n) \in \Phi_c$ une suite telle que $J_c(u_n, v_n) \to \sup_{\Phi_c} J_c$. D'après le point (2) de la proposition précédente, on peut supposer que u_n est c−concave, et que $v_n = u_n^c$. Posons $\psi_n = u_n - u_n(x_o)$ et $\varphi_n = v_n + u_n(x_o)$. On a $J_c(\psi_n, \varphi_n) = J_c(u_n, v_n)$, et donc la suite $(\psi_n, \varphi_n) \in \Phi_c$ est encore minimisante. D'après le point (1) de la proposition précédente, les fonctions ψ_n sont L-Lispchitz, et $|\psi_n(y)| \le |\psi_n(y) - \psi_n(y_o)| \le L^2$. De même les fonctions φ_n sont L-Lispchitz, et on a

$$|\varphi_n(x)| \le \inf_{y\in M}\{c(x,y) + |u_n(y) - u_n(x_o)|\} \le \frac{1}{2}L^2 + L^2.$$

D'après le théorème d'Ascoli, on peut extraire de la suite (ψ_n, φ_n) une suite convergeant uniformément vers un couple $(u,v) \in \Phi^c$. En passant à la limite le long de cette sous suite, on obtient que $J_c(u,v) = \sup_{\Phi_c} J_c$. En utilisant le point (2) de la proposition précédente, on peut remplacer (u,v) par (ψ, ψ^c), avec $\psi = v^c$ qui est bien c-concave. □

Le théorème précédent va permettre de donner une première information sur le support des couplages optimaux :

Proposition 3.4. *Si π^* est un couplage optimal, alors*

$$\pi^*\{(x,y); \psi(x) + \psi^c(y) = c(x,y)\} = 1.$$

Autrement dit $\operatorname{supp}\pi^* \subset \{(x,y); \psi(x) + \psi^c(y) = c(x,y)\}$.

Démonstration. Soit (ψ, ψ^c), avec ψ c-concave maximisant J_c sur Φ_c et π^* un couplage optimal. D'après la dualité de Kantorovich,

$$\int \psi(x)\,\mu(dx) + \int \psi^c(y)\,\nu(dy) = \iint c(x,y)\,\pi^*(dxdy),$$

et donc $\iint F(x,y)\,\pi^*(dxdy) = 0$, avec $F(x,y) = c(x,y) - \psi(x) - \psi^c(y) \ge 0$ et continue. On en déduit facilement que $\operatorname{supp}\pi^* \subset \{F = 0\}$ □

On admet la généralisation suivante du théorème de Rademacher :

Théorème 3.5. *Si f est une fonction localement Lipschitz sur M, alors f est différentiable en tout point de $M \setminus N$, où N est un ensemble de volume nul.*

Par différentiable en $x \in M$, on entend qu'il existe un vecteur $v \in T_xM$ tel que pour tout chemin $\mathcal{C}^1$ $\alpha(t) \in M$ tel que $\alpha(0) = x$ on a

$$f(\alpha(t)) = f(x) + t\langle v, \dot\alpha(0)\rangle + o(t).$$

Dans ce cas, on note $v = \nabla f(x)$.

On dira que f est sur-différentiable s'il existe $v \in T_xM$ tel que

$$f(\alpha(t)) \le f(x) + t\langle v, \dot\alpha(0)\rangle + o(t).$$

Un tel vecteur v est appelé un sur-gradient de f.

On dira que f est sous-différentiable s'il existe $v \in T_x M$ tel que

$$f(\alpha(t)) \geq f(x) + t\langle v, \dot{\alpha}(0)\rangle + o(t).$$

Un tel vecteur v est appelé un sous-gradient de f.
Les sous / sur - gradient d'une fonction ne sont pas uniques en général. Il est facile de voir qu'une fonction est différentiable si et seulement si elle est sur-différentiable et sous-différentiable.

Lemme 3.6. *Soit ψ une fonction c-concave.*

1. *Pour tout $x \in M$, l'ensemble $\partial^c \psi(x) = \{y \in M; \psi(x) + \psi^c(y) = c(x,y)\}$ est un fermé non vide.*
2. *Si ψ est différentiable en x, et $y \in \partial_c \psi(x)$, alors la fonction $z \mapsto c(z,y)$ est sous différentiable en x, et de plus, $\nabla\psi(x)$ est un sous-gradient de $c(\cdot, y)$ en x.*

Démonstration. (1) Comme ψ est c-concave, on a $\psi^{cc} = \psi$. Autrement dit

$$\psi(x) = \inf\{c(x,y) - \psi^c(y)\}.$$

Pour tout $x \in M$, la fonction $y \mapsto c(x,y) - \psi^c(y)$ est continue sur le compact M et atteint donc sa borne inférieure, ce qui prouve que $\partial^c \psi(x) \neq \emptyset$.

(2) Pour tout $z \in M$, on a toujours $\psi(z) + \psi^c(y) \leq c(z,y)$. Si α est un chemin $\mathcal{C}^1$ tel que $\alpha(0) = x$, alors

$$\begin{aligned} c(\alpha(t), y) &\geq \psi(\alpha(t)) + \psi^c(y) \\ &= \psi(x) + \psi^c(y) + t\langle \nabla\psi(x), \dot{\alpha}(0)\rangle + o(t) \\ &= c(x,y) + t\langle \nabla\psi(x), \dot{\alpha}(0)\rangle + o(t), \end{aligned}$$

ce qui prouve la sous-différentiabilité de la fonction au point x. □

Pour toute courbe γ de classe $\mathcal{C}^1$ par morceaux, on pose :

$$\mathcal{A}(\gamma) = \int_0^1 |\dot{\gamma}(t)|^2 \, dt.$$

On a toujours

$$\mathcal{A}(\gamma) \geq \left(\int_0^1 |\dot{\gamma}(t)| \, dt \right)^2 \geq d^2(\gamma(0), \gamma(1)).$$

De plus, si γ est une géodésique minimisante entre x_0 et x_1, on a $|\dot{\gamma}(t)| = d(x_0, x_1)$ pour tout t, et donc $\mathcal{A}(\gamma) = d^2(x_0, x_1)$. Par conséquent, on a

$$d^2(x_0, x_1) = \inf_{\gamma(0)=x_0, \gamma(1)=x_1} \mathcal{A}(\gamma).$$

Proposition 3.7. *Soit $x_0 \in M$ fixé ; la fonction $f : x \mapsto c(x_o, x)$ est sur-différentiable en tout point $x_1 \in M$. Plus précisément, si γ est une géodésique minimisante entre x_0 et x_1, alors $\dot{\gamma}(1) \in T_{x_1} M$ est un sur-gradient de f en x_1.*

Remarque 3.8. *Une conséquence de la proposition précédente est que s'il existe plus d'une géodésique joignant x_0 à x_1, alors la fonction f n'est pas différentiable en x_1. Par contraposée, si f est différentiable en x_1 alors il n'y a qu'une géodésique joignant x_0 à x_1. On peut montrer que f est différentiable en x_1 si et seulement s'il existe une seule géodésique joignant x_0 à x_1. Il suffit de modifier un peu la preuve ci-dessous.*

Démonstration. Prenons $\gamma : [0,1] \to M$ une géodésique minimisante entre x_0 et x_1. Considérons $\sigma :]-\varepsilon, \varepsilon[\times[0,1] \to M$ une fonction de classe $\mathcal{C}^1$ telle que $\sigma(0,t) = \gamma(t)$, et posons

$$a(s) = \mathcal{A}(\sigma(s,\cdot)) = \int_0^1 \left|\frac{\partial\sigma}{\partial t}(s,t)\right|^2 dt.$$

En dérivant sous le signe intégral, on obtient

$$\begin{aligned}\frac{da}{ds}(s) &= \int_0^1 \frac{d}{ds}\left\langle \frac{\partial\sigma}{\partial t}(s,t), \frac{\partial\sigma}{\partial t}(s,t)\right\rangle dt\\ &= 2\int_0^1 \left\langle \frac{D}{\partial s}\frac{\partial\sigma}{\partial t}(s,t), \frac{\partial\sigma}{\partial t}(s,t)\right\rangle dt,\end{aligned}$$

la deuxième égalité venant du fait que la dérivation est compatible avec la métrique Riemanienne. Il résulte de la symétrie de la dérivation covariante que

$$\frac{D}{\partial s}\frac{\partial\sigma}{\partial t}(s,t) = \frac{D}{\partial t}\frac{\partial\sigma}{\partial s}(s,t).$$

Donc,

$$\begin{aligned}\frac{da}{ds}(s) &= 2\int_0^1 \left\langle \frac{D}{\partial t}\frac{\partial\sigma}{\partial s}(s,t), \frac{\partial\sigma}{\partial t}(s,t)\right\rangle dt,\\ &= 2\int_0^1 \frac{d}{dt}\left\langle \frac{\partial\sigma}{\partial s}(s,t), \frac{\partial\sigma}{\partial t}(s,t)\right\rangle dt - 2\int_0^1 \left\langle \frac{\partial\sigma}{\partial s}(s,t), \frac{D}{\partial t}\frac{\partial\sigma}{\partial t}(s,t)\right\rangle dt\\ &= 2\left\langle \frac{\partial\sigma}{\partial s}(s,1), \frac{\partial\sigma}{\partial t}(s,1)\right\rangle - 2\left\langle \frac{\partial\sigma}{\partial s}(s,0), \frac{\partial\sigma}{\partial t}(s,0)\right\rangle - 2\int_0^1 \left\langle \frac{\partial\sigma}{\partial s}(s,t), \frac{D}{\partial t}\frac{\partial\sigma}{\partial t}(s,t)\right\rangle dt.\end{aligned}$$

En prenant $s = 0$, on obtient donc

$$\frac{da}{ds}(0) = 2\left\langle \frac{\partial\sigma}{\partial s}(0,1), \dot\gamma(1)\right\rangle - 2\left\langle \frac{\partial\sigma}{\partial s}(0,0), \dot\gamma(0)\right\rangle.$$

A présent prenons un chemin α de classe $\mathcal{C}^1$ défini au voisinage de 0 tel que $\alpha(0) = x_1$. Prenons n'importe qu'elle fonction $\sigma :]-\varepsilon, \varepsilon[\times[0,1] \to M$ de classe $\mathcal{C}^1$ telle que $\sigma(0,t) = \gamma(t)$, $\sigma(s,0) = x_0$, $\sigma(s,1) = \alpha(s)$. Il est clair qu'une telle fonction existe. On a alors,

$$d^2(x_0, \alpha(s)) \le a(s) = a(0) + sa'(0) + o(s) = d^2(x_0, x_1) + 2s\langle \dot\gamma(1), \dot\alpha(0)\rangle + o(s),$$

ce qui prouve que f est sur-différentiable au point x_1. □

Nous pouvons maintenant prouver le théorème de Mc Cann :

Théorème 3.9. *Soit μ une probabilité absolument continue par rapport à la mesure de volume sur M ; pour toute probabilité ν sur M telle $\mathcal{T}_2(\mu,\nu) < +\infty$, il existe un unique plan de transport optimal $\pi^* \in P(\mu,\nu)$. De plus, il existe une fonction c-concave $\psi : M \to \mathbb{R}$ telle que*

$$\pi^*\{(x,y); y = \exp_x(-\nabla\psi(x))\} = 1.$$

L'application $T(x) = \exp_x(-\nabla f(x))$ (définie pour μ presque tout x) est donc une solution du problème de Monge :

$$\int d^2(x, T(x))\,\mu(dx) = \inf_{S \text{ t.q } S\#\mu=\nu} \int d^2(x, S(x))\,\mu(dx).$$

Toute autre solution S coïncide avec T μ presque sûrement. De plus, pour μ presque tout $x \in M$, le chemin $\gamma(t) = \exp_x(-t\nabla\psi(x))$ $t \in [0,1]$ est l'unique géodésique minimisante joignant x à $T(x)$.

Démonstration. On a vu qu'il existait ψ c-concave telle que tout couplage optimal π^* vérifie

$$\pi^*(\cup_x \{x\} \times \partial^c \psi(x)) = 1.$$

La fonction ψ est Lispchitz et est donc différentiable en tout point $x \in M \setminus N$ avec $\mathrm{vol}(N) = 0$. On a $\pi^*(N \times M) = \mu(N) = 0$, et donc

$$\pi^*(\cup_{x \in M \setminus N} \{x\} \times \partial^c \psi(x)) = 1.$$

Or si $x \in M \setminus N$ et $y \in \partial^c \psi(x)$, on a vu que la fonction $z \mapsto c(z,y)$ était sous-différentiable en x de sous gradient $\nabla\psi(x)$. On vient de voir que cette fonction était toujours sur-différentiable. Elle est donc différentiable en x. En particulier, il n'existe qu'une seule géodésique minimisante σ joignant y à x. En identifiant le sur-gradient et le sous-gradient, on obtient

$$\nabla\psi(x) = \dot\sigma(1),$$

où $\sigma(0) = y$ et $\sigma(1) = x$. Le chemin $\gamma : t \mapsto \sigma(1-t)$ est une géodésique allant de x à y et partant avec une vitesse initiale $-\dot\sigma(1) = -\nabla\psi(x)$, donc $\gamma(t) = \exp_x(-t\nabla\psi(x))$ et $y = \exp_x(-\nabla\psi(x))$. □

4 Inégalités de Prekopa-Leindler sur une variété

Si $x, y \in M$, pour tout $s \in]0,1[$, on notera $Z_s(x,y)$ l'ensemble des points z tels que

$$d(x,z) = s d(x,y) \quad \text{et} \quad d(z,y) = (1-s) d(x,y).$$

On voit sans peine que

$$Z_s(x,y) = \{\gamma(s); \gamma \text{ géodésique minimisante de } x \text{ à } y\}$$

Le résultat suivant est du à Cordero-Erausquin-Mc Cann-Schmuckenschläger.

Théorème 4.1. *Soit μ une probabilité sur M ayant une densité de la forme $\mu(dx) = e^{-V(x)}\mathrm{vol}(dx)$, avec V une fonction de classe $\mathcal{C}^2$. On suppose que μ vérifie la condition de courbure $\mathrm{CD}(K,\infty)$ $K \in \mathbb{R}$ de Bakry-Emery, c'est à dire que pour tout $x \in M$,*

$$\mathrm{Ric}_x(v,v) + \mathrm{Hess}_x(v,v) \geq K|v|_x^2, \qquad \forall v \in T_xM.$$

Si $f, g, h : M \to \mathbb{R}^+$ sont trois fonctions vérifiant pour un certain $s \in [0,1]$

$$h(z) \geq e^{-Ks(1-s)d^2(x,y)/2} f(x)^s g(y)^{1-s}, \qquad \forall x, y \in M, \forall z \in Z_s(x,y),$$

alors

$$\int h \, d\mu \geq \left(\int f \, d\mu \right)^s \left(\int g \, d\mu \right)^{1-s}.$$

Démonstration. On peut supposer que f et g sont de masse 1 sous μ ; on cherche à démontrer que $\int h \, d\mu \geq 1$. Considérons l'application de transport T qui envoie $\mu_0 = f\, d\mu$ sur $\mu_1 = g\, d\mu$. Il existe ψ $d^2/2$ concave telle que

$$T(x) = \exp_x(-\nabla\psi(x)).$$

On pose pour tout $t \in [0,1]$ $T_t(x) = \exp_x(-t\nabla\psi(x))$. Il résulte du théorème de Mc Cann que $T_s(x) \in Z_s(x, T(x))$ pour μ_0 presque tout x.
Grâce à un changement de variable que nous ne justifierons pas, on obtient

$$\begin{aligned}\int h\, d\mu &= \int h(T_s(x))|\det(dT_s)(x)|e^{-V(T_s(x))}\,\mathrm{vol}(dx)\\ &\geq \int e^{-Ks(1-s)d^2(x,T(x))/2} f(x)^s g(T(x))^{1-s}|\det(dT_s)(x)|e^{-V(T_s(x))}\,\mathrm{vol}(dx).\end{aligned}$$

Grâce à un autre changement de variable, on obtient l'équation de Monge-Ampère :

$$f(x)e^{-V(x)} = g(T(x))e^{-V(T(x))}|\det(dT)(x)|.$$

Posons $\alpha_s(x) = -\log|\det(dT_s(x))|$. Pour $s = 0$, on a $T_0 = \mathrm{Id}$ et donc $\alpha_0 = 0$.

$$\int h\, d\mu \geq \int e^{-Ks(1-s)d^2(x,T(x))/2} e^{(1-s)V(T(x))+sV(x)-V(T_s(x))} e^{-\alpha_s(x)+(1-s)\alpha_1(x)} f(x)\,\mu(dx).$$

Il suffit de montrer que pour μ presque tout x,

$$V(T_s(x)) + \alpha_s(x) \leq (1-s)\left[V(T_1(x)) + \alpha_1(x)\right] + s\left[V(T_0(x)) + \alpha_0(x)\right] - Ks(1-s)d^2(x,T(x))/2$$

et pour cela, il suffit de montrer que

$$\forall x \in M, \qquad \forall u \in [0,1], \qquad \frac{d^2}{du^2}(V(T_u(x)) + \alpha_u(x)) \geq Kd^2(x, T(x)).$$

Observons que $u \mapsto T_u(x)$ est une géodésique minimisante joignant x à $T(x)$. On a donc

$$\frac{d^2}{du^2}(V(T_u(x))) = \mathrm{Hess}_{T_u(x)}\left(\frac{d}{du}T_u(x), \frac{d}{du}T_u(x)\right)$$

et d'après le lemme 4.2 suivant

$$\frac{d^2}{du^2}\alpha_u(x) \geq \mathrm{Ric}\left(\frac{d}{du}T_u(x), \frac{d}{du}T_u(x)\right).$$

Donc

$$\frac{d^2}{du^2}(V(T_u(x)) + \alpha_u(x)) \geq K\left|\frac{d}{du}T_u(x)\right|^2 = Kd^2(x, T(x)).$$

□

Lemme 4.2.

$$\frac{d^2}{du^2}\alpha_u(x) - \frac{1}{n}\left(\frac{d}{du}\alpha_u(x)\right)^2 - \mathrm{Ric}\left(\frac{d}{du}T_u(x), \frac{d}{du}T_u(x)\right) \geq 0.$$

La preuve de ce lemme sera donnée lors de la prochaine séance du groupe de travail.

A suivre...