Chapitre 11

Transport optimal

11.1 Introduction

Nous avons déjà observé plusieurs manières d'aborder la concentration de la mesure : via des résultats isopérimétriques mais aussi grâce à des inégalités fonctionnelles. Ce chapitre à pour objectif d'introduire un nouveau point de vue, celui du transport optimal. Comme nous le verrons, cette théorie permet facilement d'obtenir des résultats de concentrations mais il met également en avant de nouvelles quantités comme les distances de Wasserstein qui sont utiles pour quantifier la convergence de mesure. Dans ce qui suit, nous allons commencer par énoncer le problème de Monge-Kantorovich et présenter les nouveaux objets qui lui sont liés. Nous aborderons ensuite certains aspects qualitatifs et quantitatifs de ce problème pour enfin présenter les liens existant entre transport optimal et concentration de la mesure.

11.1.1 Problème de Monge-Kantorovich

Dans ce qui suit (E, d) désigne un espace polonais et rappelons que $\mathcal{P}(E)$ représente l'ensemble des mesures de probabilités sur les boréliens de E.

Définition 11.1.1. *Soient μ et ν deux mesures de probabilités sur E. Une application $T : E \to E$ transporte μ sur ν si*

$$\int_E f(y)d\nu(y) = \int_E f(T(x))d\mu(x)$$

pour toute fonction $f : E \to \mathbb{R}$ borélienne bornée.

Définition 11.1.2. *Soient $\mu, \nu \in \mathcal{P}(E)$. Un couplage entre μ et ν est une mesure de probabilité π sur $E \times E$ dont la première marginale vaut μ et la seconde ν. Nous désignerons par $\Pi(\mu, \nu)$ l'ensemble de tels couplages*

Remarque. D'un point de vu probabiliste, un couplage n'est autre que la loi d'un couple de variable aléatoire (X, Y) telles que $\mathcal{L}(X) = \mu$ et $\mathcal{L}(Y) = \nu$.

Exemple 11.1.1. 1. Le couplage le plus simple et celui qui consiste à choisir $\pi = \mu \otimes \nu$. Autrement dit, $\pi = \mathcal{L}\big((X, Y)\big)$ avec X et Y des variables aléatoires indépendantes de loi respective μ et ν.

2. Etant donné une application $T : E \to E$ transportant μ sur ν, il est possible de produire un couplage déterministe en posant $\pi_T = \mathcal{L}(X, T(X))$. En particulier,

$$\iint_{E\times E} f(x,y)d\pi_T(x,y) = \iint_{E\times E} f\big(x, T(x)\big)d\mu(x)$$

pour toute fonction $f : E \to \mathbb{R}$ borélienne bornée.

Maintenant que nous avons introduit quelques points de vocabulaire, nous pouvons énoncer le problème de Monge-Kantorovich.

Définition 11.1.3. *Etant donnés deux mesures de probabilités $\mu, \nu \in \mathcal{P}(E)$ ainsi qu'une fonction mesurable $c : E \times E \to [0, +\infty]$, le problème de Monge-Kantorovich consiste à minimiser le coût de transport (entre μ et ν) défini par*

$$\mathfrak{I}_c(\pi) = \iint_{E\times E} c(x,y)d\pi(x,y)$$

sous la contrainte $\pi \in \Pi(\mu, \nu)$.

Le coût de transport optimal entre μ et ν sera noté par

$$\mathcal{T}_c(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \mathfrak{I}_c(\pi).$$

11.1.2 Problème de Monge-Kantorovich, résultats qualitatifs

Lorsque la fonction de coût c satisfait certaines hypothèses, le théorème suivant nous assure que le problème de Monge-Kantorovich admet une solution.

Théorème 11.1.1. *Si $c : E \times E \to [0, +\infty]$ est une fonction semi-continue inférieurement alors, pour toutes mesures $\mu, \nu \in \mathcal{P}(E)$, il existe $\pi^* \in \Pi(\mu, \nu)$ telle que*

$$\mathcal{T}_c(\mu,\nu) = \iint_{E\times E} c(x,y)\pi^*(x,y)$$

Remarque. π^* porte le nom de *plan de transport optimal* associé aux mesures μ et ν.

Démonstration. Voici les grandes lignes de la démonstration.

1. Tout d'abord, il faut montrer que $\mathfrak{I}_c$ est une application semi-continue inférieurement sur $\mathcal{P}(E \times E)$.

2. L'ensemble $\Pi(\mu, \nu)$ est ensemble compact pour la topologie de la convergence étroite. Pour démontrer ceci, il faudra faire appel au Théorème de Prokhorov qui caractérise la compacité relative d'un ensemble de mesure en terme de tension.

3. Pour conclure, il suffit d'utiliser le fait suivant : une fonction semi-continue inférieurement sur un ensemble compact est minorée et atteint sa borne inférieure.

Le lecteur est prié de se référer aux ouvrages suivants pour une démonstration de ce théorème [144, 143]. □

En dimension un, il est possible de préciser la forme du couplage optimal à l'aide des fonctions de repartition. Si $\mu \in \mathcal{P}(\mathbb{R})$, nous noterons sa fonction de répartition par $F_\mu(x) = \mu\big(]-\infty, x]\big)$.

Proposition 11.1.1. *Soient $\mu, \nu \in \mathcal{P}(E)$. Si μ une mesure sans atome alors*

$$\pi^* = \mathcal{L}\big(X, T(X)\big) \quad \text{où} \quad T(x) = F_\nu^{-1} \circ F_\mu(x)$$

avec $\mathcal{L}(X) = \mu$. De plus, T est une application croissante.

Remarque. Dire que μ est une mesure sans atome signifie que $\mu(a) = 0$ pour tout $a \in \mathbb{R}$. Il aurait également pu être envisageable de choisir $\pi^* = \mathcal{L}\big(X, S(X)\big)$ avec

$$S(x) = F_\nu^{-1} \circ (1 - F_\mu(x)) \quad x \in \mathbb{R}$$

qui est une application décroissante.

11.1.3 Dualité de Kantorovich

Tout comme de nombreux problème de minimisation, celui de Monge-Kantorovich peut également s'énoncer de manière duale. Comme nous le verrons, cette nouvelle formulation sera utile pour établir un lien entre la théorie du transport optimal et la concentration de la mesure ; ceci s'effectuera par le biais de nouvelles inégalités fonctionnelles.

Théorème 11.1.2 (Dualité)**.** *Soient $c : E \times E \to [0 + \infty[$ une fonction de coût semi-continue inférieurement et $\mu, \nu \in \mathcal{P}(E)$ telles que $\mathfrak{I}_c(\mu, \nu) < +\infty$. Alors*

$$\mathcal{T}_c(\mu, \nu) = \sup_{(\phi,\psi)\in\Phi_c} \left\{ \int_E \psi(x)\mu(dx) + \int_E \phi(y)\nu(dy) \right\} \tag{11.1.1}$$

où $\Phi_c = \left\{ (\psi, \phi) \in C_b^0(E) \times C_b^0(E) \,;\, \psi(x) + \phi(y) \le c(x, y) \quad \textit{pour} \quad x, y \in E \right\}$.

Remarque.

1. Il est possible de relâcher la condition sur le couple (ψ, ϕ) en supposant que $\psi \in L^1(\mu)$ et $\phi \in L^1(\nu)$.

2. La démonstration permettant d'établir cette dualité utilise la notion de c-convexité et de monotonie cyclique par rapport à la fonction de coût c. Plus précisément, il y a équivalence entre les assertions suivantes :

   - π est un couplage optimal ;
   - le support de π est un ensemble c-cycliquement monotone ;
   - il existe une fonction f c-convexe.

   Toutefois pour ne pas alourdir le cours, nous ne développerons pas ces aspects et renvoyons le lecteur vers [144, 143]. Néanmoins, nous reviendrons brièvement sur cette notion de c-convexité lorsque nous aborderons les opérateurs d'inf-convolution permettant de définir le semi-groupe d'Hamilton-Jacobi.

11.1.4 Théorème de Brenier

Lorsque $E = \mathbb{R}^n$ et $c(x, y) = \frac{1}{2}|x - y|^2$ (avec $|\cdot|$ la distance euclidienne), la solution problème de Monge-Kantorovich admet une forme élégante que nous allons présenter ci-dessous. Pour un tel choix de fonction de coût, nous adopterons la notation suivante

$$\mathcal{T}_2(\mu, \nu) = \mathcal{T}_c(\mu, \nu) \quad \text{avec} \quad \mu, \nu \in \mathcal{P}(\mathbb{R}^n).$$

Théorème 11.1.3 (Brenier)**.** *Soient μ, ν deux mesures de probabilités sur $(\mathbb{R}^n, |\cdot|)$; si μ est absolument continue par rapport à la mesure de Lebesgue et si $\mathcal{T}_2(\mu, \nu) < +\infty$ alors*

1. *il existe un unique plan de transport optimal $\pi^* \in \Pi(\mu, \nu)$ et celui-ci est déterministe*

$$i.e. \quad \pi^* = \mathcal{L}(X, T(X)) \quad pour \quad T : \mathbb{R}^n \to \mathbb{R}^n$$

   avec $\mathcal{L}(X) = \mu$.

2. *il existe une fonction convexe $\phi : \mathbb{R}^n \to \mathbb{R}$ (finie μ-presque partout) telle que $T = \nabla\phi$.*

3. *l'application T est essentiellement unique : si $\tilde{T}$ définie une autre solution optimale alors $\tilde{T} = T$ μ-presque partout.*

Remarque. 1. Ce résultat a été étendu aux cadres des variétés riemaniennes par McCann. L'expression de l'application T est un peu plus complexe et fait intervenir la fonction exponentielle associée à la variété.

2. En pratique, le Théorème de Brenier met en jeu un changement de variable qui s'apparente à une équation de Monge-Ampère. Nous reviendrons ce sur point dans une section ultérieure.

Démonstration. La démonstration de ce résultat est délicate et sera omise dans ce cours. Mentionnons toutefois, qu'elle fait appel à des résultats d'analyse convexe (notion de sous-différentielle, différentiabilité presque partout d'une fonction convexe,. . .) mais aussi à la caractérisation des plans de transports optimaux en terme de c-convexité et de c-monotonie cyclique pour le coût quadratique euclidien. Le lecteur trouvera la démonstration de ce résultats dans [144, 143]. □

11.1.5 Equation de Monge Ampère

Comme mentionné un peu plus haut, lorsque $c(x, y) = \frac{1}{2}|x - y|^2$, le Théorème de Brenier donne lieu à un changement de variable. En effet, sous les hypothèses du Théorème de Brenier, il existe une application $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ telle que l'application T transportant μ sur ν soit de la forme

$$T = \nabla\phi.$$

En particulier, si $d\mu = gdx$ et $d\nu = fdx$, nous avons (par définition de T)

$$\int_{\mathbb{R}^n} h(y)g(y)dy = \int_{\mathbb{R}^n} h\big(T(x)\big)f(x)dx$$

pour toute fonction borélienne bornée $h : \mathbb{R}^n \to \mathbb{R}$. Lorsque T est un C^1-difféomorphisme, il est alors possible de procéder au changement de variable $y = T(x)$ dans la première intégrale. Ainsi, nous obtenons

$$\int_{\mathbb{R}^n} h\big(T(x)\big)g\big(T(x)\big)\big|\det \mathrm{Jac}\, T(x)\big|dx = \int_{\mathbb{R}^n} h\big(T(x)\big)f(x)dx$$

où $\mathrm{Jac}\, T(x)$ désigne la matrice jacobienne de T au point x. En conséquence de ceci, pour tout $x \in \mathbb{R}^n$, nous avons

$$f(x) = g\big(T(x)\big)\big|\det \mathrm{Jac}\, T(x)\big|.$$

En outre, puisque $T = \nabla\phi$, ceci s'écrit

$$f(x) = g\big(\nabla\phi(x)\big)\big|\det \mathrm{Hess}\,\phi(x)\big| \tag{11.1.2}$$

où $\mathrm{Hess}\,\phi(x)$ désigne la matrice (semi-définie positive) de ϕ au point x. L'équation (11.1.2) est désignée sous le nom d'équation de Monge-Ampère (cf. [144, 143] pour plus de détails). Comme nous le verrons ultérieurement, il est possible d'obtenir de nombreuses informations à partir de cette équation.

Notons que les hypothèses sur T faites ci-dessus sont trop restrictives et irréalistes. En effet, puisque ϕ est convexe, $\nabla\phi$ est définie presque partout mais il n'y a, a priori, aucune raison pour que $\nabla\phi$ soit bijective ou régulière. Toutefois, il est possible d'affaiblir les hypothèses faites sur T pour donner un sens à l'équation de Monge-Ampère (11.1.2) dans un contexte plus général. Cette généralisation s'effectue via la notion de sous-gradient qui permet de définir une Hessienne au sens d'Alexandrov. Ces résultats plus complexes ne seront pas abordés dans ce cours.

11.2 Problème de Monge-Kantorovich, aspect métrique

Rappelons que l'espace sous-jacent E est un espace polonais, en particulier E est muni d'une distance d. Lorsque la fonction de coût c coïncide avec la distance d, cela donne lieu a des quantités interessantes permettant de définir une nouvelle distance sur l'espace des mesures de probabilités sur E (en imposant quelques conditions de moments).

D'un point de vue heuristique, la quantité $\mathcal{T}_c(\mu,\nu)$ permet de mesurer la différence entre deux mesures de probabilités. C'est pourquoi, intuitivement, si $\mu = \nu$, il est naturel de pressentir que le meilleur moyen de minimiser la fonctionnelle de coût est de ne rien faire : autrement dit $\mathcal{T}_c(\mu,\mu) = 0$. La propriété de symétrie de l'application $\mathcal{I}_c(\mu,\nu)$ est, quant à elle, évident ; il parait alors réaliste d'espérer qu'une inégalité triangulaire soit satisfaite.

Avant d'approfondir et de développer les quelques lignes qui précèdent, nous allons introduire de nouvelles notations. Pour $p \in [1, +\infty[$ et deux mesures de probabilités $\mu, \nu \in \mathcal{P}(E)$, posons

$$\mathcal{T}_p(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \iint_{E\times E} d(x,y)^p d\pi(x,y)$$

Observons également le fait suivant : si $x_0 \in E$ et $\mu \in \mathcal{P}(E)$ alors

$$\mathcal{T}_p(\delta_{x_0},\mu) = \int_E d(x_0,y)^p d\mu(y)$$

Suite à cette observation, nous allons nous restreindre aux mesures pour lesquelles le coût précédent est fini.

Définition 11.2.1. *Pour $p \in [1, +\infty[$, nous désignons par $\mathcal{P}_p(E)$ l'ensemble des mesures boréliennes sur E qui admettent un moment d'ordre p :*

$$\mathcal{P}_p(E) = \left\{ \mu \in \mathcal{P}(E) \, ; \, \int_E d(x_0, x)^p d\mu(x) < +\infty \right\}$$

pour un certain $x_0 \in E$.

Remarque. Il est élémentaire de vérifier que $\mathcal{P}_p(E)$ ne dépend pas du choix de x_0.

Il est maintenant possible de définir une distance sur $\mathcal{P}_p(E)$ à l'aide de $\mathcal{T}_p$. Ces nouvelles distances portent le nom de distance de Kantorovich-Wasserstein.

Proposition 11.2.1. *La quantité*

$$W_p(\mu, \nu) = \mathcal{T}_p(\mu, \nu)^{1/p}$$

définie pour tout $\mu, \nu \in \mathcal{P}_p(E)$ est une distance sur $\mathcal{P}_p(E)$.

Remarque. En particulier, si $\mu = \delta_x$ et $\nu = \delta_y$ alors $W_p(\mu, \nu) = d(x, y)$.

Démonstration. La vérification de l'inégalité triangulaire est laissée en exercice. □

Comme pour les normes associées $\| \cdot \|_p$ aux espaces $L^p(\mu)$, il est possible de comparer les distances de Kantorovich-Wassertein entre elles.

Proposition 11.2.2. *Pour $1 \leq p \leq q$ et pour toutes mesures $\mu, \nu \in \mathcal{P}_q$ nous avons*

$$W_p(\mu, \nu) \leq W_q(\mu, \nu).$$

Démonstration. Soient $1 \leq p \leq q$ et rappelons que pour toutes mesures de probabilité μ, $L^q(\mu) \subset L^p(\mu)$, c'est pourquoi $\mathcal{P}_q \subset \mathcal{P}_p$. Si (X, Y), avec $\mathcal{L}(X) = \mu$ et $\mathcal{L}(Y) = \nu$, désigne un couple de variables aléatoires qui réalise $W_q(\mu, \nu)$ alors

$$\big\| d(X, Y) \big\|_p \leq \big\| d(X, Y) \big\|_q = W_q(\mu, \nu).$$

Or, par definition de $W_p(\mu, \nu)$ nous avons aussi

$$W_p(\mu, \nu) \leq \big\| d(X, Y) \big\|_p$$

d'où le résultat. □

Caractérisation duale de W_p

Comme nous allons le voir, il est possible de caractériser de manière duale les distances $(W_p)_{p \geq 0}$. Nous allons nous focaliser sur les cas $p = 1$ et $p = 2$.

Théorème 11.2.1 (Dualité pour W_1). *Pour tout $\mu, \nu \in \mathcal{P}_1$, la distance $W_1(\mu,\nu)$ est donnée par*

$$W_1(\mu,\nu) = \sup_{f\in\mathcal{F}_1} \left\{ \left| \int_E f d\mu - \int_E f d\nu \right| \right\} \tag{11.2.1}$$

avec $\mathcal{F}_1$ désignant les fonctions 1-lipschitziennes.

Démonstration. Désignons par S le précédent supremum et montrons que $S = W_1(\mu,\nu)$. A cet effet, soient f une fonction 1-lipschitzienne et (X,Y) un couple de variables aléatoires tel que $\mathcal{L}\big((X,Y)\big) = \pi$ (i.e. $\pi \in \Pi(\mu,\nu)$). Alors

$$\begin{aligned} \left| \int_E f d\mu - \int_E f d\nu \right| &= \big|\mathbb{E}_\mu[f(X)] - \mathbb{E}_\nu[f(Y)]\big| \\ &\le \mathbb{E}_\pi\big[\big|f(X) - f(Y)\big|\big] \\ &\le \mathbb{E}_\pi\big[\big|d(X,Y)\big|\big] \end{aligned}$$

puisque f est 1-lipschitzienne. Ainsi, $S \le \mathbb{E}_\pi[|d(X,Y)|]$ pour tout couplage $\pi \in \Pi(\mu,\nu)$; c'est pourquoi $S \le W_1(\mu,\nu)$.

Démontrons à présent que $S \ge W_1(\mu,\nu)$. Soit $\epsilon > 0$ fixé, d'après la dualité de Kantorovich (i.e. le théorème 11.1.2), il existe $\psi \in L^1(\mu)$ et $\phi \in L^1(\nu)$ telles que, pour tout $x, y \in E$

$$\psi(x) + \phi(y) \le d(x,y) \quad \text{et} \quad \int_E \psi d\mu + \int_E \phi d\nu \ge W_1(\mu,\nu) - \epsilon. \tag{11.2.2}$$

Posons alors $f(x) = \sup_{y\in E}\big(\phi(y) - d(x,y)\big)$. Il n'est pas difficile de montrer, via l'inégalité triangulaire, que f est 1-lipschitzienne et que $f \in L^1(\mu) \cap L^1(\nu)$ puisque $\mu, \nu \in \mathcal{P}_1(E)$.

Observons de plus, en choisissant $y = x$ dans le supremum définissant f, que $f(x) \ge \phi(x)$. En outre, par hypothèse nous avons $\psi(x) + \phi(y) \le d(x,y)$, ce qui permet d'obtenir l'inégalité suivante

$$f(x) = \sup_{y\in E}\big(\psi(y) - d(x,y)\big) \le -\psi(x) \quad \text{avec} \quad x \in E. \tag{11.2.3}$$

En combinant les inégalités (11.2.2) et (11.2.3) nous obtenons

$$S \ge -\int_E f d\mu + \int_E f d\nu \ge \int_E \psi d\mu + \int_E \phi d\nu \ge W_1(\mu,\nu) - \epsilon$$

ce qui achève la démonstration. □

Les mêmes arguments permettent d'obtenir un résultat similaire lorsque $p = 2$.

Théorème 11.2.2 (Dualité pour W_2). *Pour tout $\mu, \nu \in \mathcal{P}_2$, la distance $W_2(\mu,\nu)$ est donnée par*

$$W_2(\mu,\nu) = \sup_{f\in\mathcal{F}_2} \left\{ \left| \int_E \psi d\mu - \int_E \phi d\nu \right| \right\} \tag{11.2.4}$$

avec $\mathcal{F}_2$ désignant l'ensemble des fonctions bornées ψ et ϕ telles que

$$\psi(x) \le \phi(y) + \frac{1}{2}|x-y|^2 \quad \textit{avec} \quad x, y \in E.$$

Remarque. Le choix optimal, parmi les fonctions appartenant à $\mathcal{F}_2$, est obtenu en choisissant ψ de la manière suivante

$$\psi(x) = \inf_{y\in\mathbb{R}^n} \big[\phi(y) + \frac{1}{2}|x-y|^2\big].$$

ψ est usuellement désignée comme l'infimum-convolution de la fonction ϕ pour le coût quadratique $c(x,y) = \frac{1}{2}|x-y|^2$.

11.2.1 Lien avec la topologie faible

Les distances $(W_p)_{p\geq 1}$ peuvent être utilisées en rapport avec la convergence faible. Plus précisément,

Théorème 11.2.3. *Soient μ et $(\mu_n)_{n\geq 1}$ des mesures dans $\mathcal{P}_p(E)$. Dans ce cas, $W_p(\mu_n,\mu)$ tend vers 0 lorsque $n\to+\infty$ si et seulement si*

1. *$(\mu_n)_{n\geq 1}$ converge faiblement vers μ ;*
2. *les moments d'ordre p de $(\mu_n)_{n\geq 1}$ convergent vers ceux de μ :*

$$\lim_{n\to+\infty} \int_E d(x_0,x)^p d\mu_n(x) = \int_E d(x_0,x)^p d\mu(x).$$

Démonstration. Nous admettrons cette preuve qui nous éloignerait de notre sujet. Notons toutefois que le sens direct de la démonstration utilise le Théorème du portemanteau (pour la convergence faible) et l'inégalité suivante : pour tout $\epsilon > 0$, tout $p \geq 1$, il existe $C(\epsilon, p)$ telle que

$$|a^p - b^p| \leq \epsilon a^p + C(\epsilon,p)|b-a|^p \quad \text{pour tout} \quad (a,b)\in\mathbb{R}_+^2$$

pour la convergence des moments. Le sens réciproque utilise des résultats d'uniforme intégrabilité ainsi que l'équivalence entre convergence en loi et convergence en probabilité d'une suite de variables aléatoires lorsque la limite est une constante. □

11.2.2 Applications

Maintenant que nous avons préciser de quelle manière les distances de Kantorovich-Wasserstein sont liées à la convergence de mesure, voyons quelques illustrations de ce résultat.

Théorème de la limite centrale

Débutons par un résultat classique. Soit X une variable aléatoire, centrée, réduite, de loi μ et de carré intégrable et rappelons que γ_1 désigne la mesure gaussienne standard dans $\mathbb{R}$. Considérons $(X_i)_{i\geq 1}$ une suite de copies indépendantes de X. Comme auparavant, $S_n = \sum_{i=1}^n X_i$.

Théorème 11.2.4 (Théorème de la limite centrale)**.** *Dans le cadre précédent, nous avons*

$$\lim_{n\to+\infty} W_2^2(\mu_n,\gamma_1) = 0 \tag{11.2.5}$$

où $\mathcal{L}(n^{-1/2}S_n) = \mu_n$. En particulier, $n^{-1/2}S_n$ converge en loi vers Z avec $\mathcal{L}(Z) = \gamma_1$.

Etude des files d'attentes $M/M/\infty$

Il est également possible d'aborder des résultats plus complexes impliquant de la dépendance. A cet effet, rappelons qu'il est usuel de modéliser certaines files d'attentes de la manière suivante :

1. la file d'attente est composée de $x_0 \in \mathbb{R}$ personnes.
2. A chaque instant, une personne est servie avec probabilité $p \in [0,1]$ et sort de la file d'attente ; bien entendu, avec probabilité $q = 1-p$ cette personne reste dans la file.
3. de nouvelles personnes s'insèrent dans la file d'attente. Cette quantité est déterminée par une loi de Poisson de paramètre $\lambda > 0$.

Toutes les variables aléatoires mentionnées (implicitement ou non) ci-dessus, sont supposées indépendantes et nous désignons par $(X_n)_{n\geq 0}$ le nombre de personnes dans la file à l'instant $n \in \mathbb{N}$. Il se trouve qu'il s'agit d'une chaine de Markov dont il est possible de préciser certaines de ses propriétés.

Dans ce qui suit, pour tout $k, l \in \mathbb{N}$, nous noterons pas $p_{k,l}$ les probabilités de transitions,

$$i.e. \quad p_{k,l} = \mathbb{P}(X_1 = l, X_0 = k)$$

Théorème 11.2.5. *Dans le cadre précédent, nous avons*

1. *pour tout* $k, l \in \mathbb{N}$, $p_{k,l} = \sum_{j=0}^{k\wedge l} \frac{k!}{j!(k-j)!(l-j)!} q^j p^{k-j} e^{-\lambda} \lambda^{l-j}$.
2. *La chaine* $(X_n)_{n\geq 0}$ *est irréductible, apériodique, récurrente positive, de probabilité invariante* $\nu = \mathcal{P}(\frac{\lambda}{p})$.

Le théorème précédent combiné à un résultat classique de probabilité (cf. [17]), nous assure que la chaine de Markov converge en loi vers la mesure ν. Nous allons tâcher de quantifier cette vitesse de convergence à l'aide de distance la distance de Wasserstein W_1.

Théorème 11.2.6. *Pour toute loi initiale* μ_0 *admettant un moment d'ordre* 1 *et pour tout* $n \in \mathbb{N}$ *nous avons*

$$W_1(\mu_n, \nu) \leq q^n W_1(\mu_0, \nu) \tag{11.2.6}$$

avec $\mathcal{L}(X_n) = \mu_n$.

Démonstration. Avant toutes choses, introduisons des notations :

- $(B_{n,l})$ désigne une famille de variables aléatoires de Bernoulli indépendantes de paramètre q telles que $B_{n,l} = 0$ si le l-ième client est servi au temps n ; $B_{n,l} = 1$ sinon.
- $(A_n)_{n\geq 0}$ une famille de variables aléatoires de Poisson de paramètre $\lambda > 0$ telle que A_n désigne le nombre d'arrivée au temps $n \in \mathbb{N}$.

Dans ce qui suit, nous identifierons variables aléatoires avec leurs lois de probabilités. L'idée essentielle de la preuve consiste, à partir d'une loi initiale (X_0, Y_0), de construire un couplage $(X_n, Y_n)_{n\in\mathbb{N}}$. Il sera essentiel que les processus $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ soient construit à partir du même aléa. Ces processus sont définis par récurrence comme suit :

$$X_{n+1} = \sum_{k=1}^{X_n} B_{n+1,k} + A_{n+1} \quad \text{et} \quad Y_{n+1} = \sum_{k=1}^{Y_n} B_{n+1,k} + A_{n+1}.$$

Pris séparément, $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ sont des chaines $M/M/\infty$. L'avantage dans ce procédé de construction est qu'il devient aisé de calculer la différence entre ces deux processus, pris au même instant :

$$|X_{n+1} - Y_{n+1}| = \sum_{k=\min(X_n,Y_n)+1}^{\max(X_n,Y_n)} B_{n+1,k}.$$

Alors, si $\mathcal{F}_n$ désigne la tribu engendrée par les variables $(B_{k,l}, A_l)_{k\leq n,l}$, nous avons

$$\mathbb{E}\Big[\big|X_{n+1} - Y_{n+1}\big|\Big|\mathcal{F}_n\Big] = q\big|X_n - Y_n\big| \quad \text{pour tout} \quad n \in \mathbb{N}.$$

A l'aide d'une récurrence immédiate sur $n \in \mathbb{N}$, nous en déduisons que $\mathbb{E}\big[\big|X_n-Y_n\big|\big] = q^n\mathbb{E}\big[\big|X_0-Y_0\big|\big]$ pour tout $n \in \mathbb{N}$.

Pour conclure, considérons une loi initiale μ_0 admettant un moment d'ordre 1. Désignons alors un couple (X_0, Y_0) tel que $W_1(\mu_0, \nu) = \mathbb{E}\big[\big|X_0, Y_0\big|\big]$. Ainsi, par définition de W_1, nous avons

$$W_1(\mu_n, \nu) \leq \mathbb{E}\big[\big|X_n - Y_n\big|\big] \leq q^n\mathbb{E}\big[\big|X_0 - Y_0\big|\big]$$

d'où le résultat. □

11.3 Inégalité de Transport et concentration

Dans cette section, nous allons présenter les liens entre les inégalités de transport et la concentration de la mesure.

Définition 11.3.1. *Une mesure $\mu \in \mathcal{P}_p(E)$ satisfait une inégalité de transport $T_p(C)$ (avec $p \geq 1$) s'il existe une constante $C = C_p > 0$ telle que*

$$W_p \leq \sqrt{2CH(\mu,\nu)}, \quad \textit{pour tout} \quad \nu \in \mathcal{P}_p(E) \tag{11.3.1}$$

avec $H(\mu,\nu) = \int_E \log \frac{d\nu}{d\mu} d\nu$ l'entropie relative de ν par rapport à ν.

11.3.1 Transport et concentration, argument de Marton

Dans ce qui suit, nous supposons que μ vérifie une inégalité de transport $T_1(C)$. Comme observé par Marton (cf. [105]), nous allons voir qu'il est possible d'obtenir un résultat de concentration pour la mesure μ à partir de l'inégalité $T_1(C)$ (11.3.1).

En effet, considérons deux ensemble boréliens A et B et désignons par μ_A (respectivement μ_B) la mesure de probabilité induite par la restriction de μ à l'ensemble A (respectivement à l'ensemble B). En combinant l'inégalité triangulaire satisfaite par W_1 avec l'inégalité (11.3.1) (avec $p = 1$), nous obtenons

$$\begin{aligned} W_1(\mu_1,\mu_B) &\leq W_1(\mu,\mu_A)+W_1(\mu,\mu_B) \\ &\leq \sqrt{2CH(\mu_A|\mu)}+\sqrt{2CH(\mu_B|\mu)} \\ &= \sqrt{2C\log\frac{1}{\mu(A)}}+\sqrt{2C\log\frac{1}{\mu(B)}} \end{aligned}$$

où la dernière égalité provient du calcul explicite des entropies relatives. De plus, observons que toute mesure de probabilité ayant pour lois marginales μ_A et μ_B doit avoir son support dans $A\times B$. Ainsi, par définition de W_1,

$$W_1(\mu_A,\mu_B)\geq d(A,B)=\inf\big\{d(x,y)\,;\,x\in A,\,y\in B\big\}.$$

Enfin, si A et B sont choisis dans E tels que $d(A,B)\geq r>0$, par ce qui précède, nous obtenons

$$r\leq\sqrt{2C\log\frac{1}{\mu(A)}}\sqrt{2C\log\frac{1}{\mu(A_r)}} \tag{11.3.2}$$

avec, rappelons le, $A_r=\big\{x\in E\,;\,d(x,A)<r\big\}$. L'inégalité (11.3.2) permet alors d'obtenir de la concentration. En effet, si $\mu(A)\leq\frac{1}{2}$, celle-ci entraine que

$$r\leq\sqrt{2C\log 2}+\sqrt{2C\log\frac{1}{1-\mu(A_r)}}.$$

D'où, lorsque $r\geq 2\sqrt{2C\log 2}$, nous avons

$$1-\mu(A_r)\leq e^{-\frac{r^2}{8C}}.$$

Comme cela a été démontré par Bobkov et Götze (cf. [23]), il est possible de préciser l'observation de Marton. C'est le contenu du résultat suivant.

Théorème 11.3.1 ($T_1(C)$ et concentration)**.** *Soit μ une mesure de probabilité sur un espace métrique (E,d) alors*

$$W_1(\mu,\nu)\leq\sqrt{2C_1H(\nu|\mu)}\quad\textit{pour tout}\quad \nu\in\mathcal{P}_1(E) \tag{11.3.3}$$

avec $C_1>0$ une constante numérique si et seulement si

$$\sup_{F\in\mathrm{Lip}(1)}\mathbb{E}[e^{\lambda F(X)}]\leq e^{\frac{C_1\lambda^2}{2}}\quad\textit{pour tout}\quad\lambda\in\mathbb{R} \tag{11.3.4}$$

avec $\mathcal{L}(X)=\mu$ et Lip(1) *désigne l'ensemble des fonctions $F:E\to\mathbb{R}$ 1-lipschitziennes.*

Remarque. La présence de la transformée de Laplace permet de montrer que les inégalités de transport $T_1(C)$ ne sont pas adaptées pour produire des inégalités de concentration indépendantes de la dimension. Autrement dit, les inégalités de transport $T_1(C)$ ne se tensorisent pas indépendamment de la dimension

Démonstration. La démonstration repose sur la formulation duale de W_1 (celle donnée par le théorème 11.1.2). Celle-ci affirme que

$$W_1(\mu,\nu) = \sup\left[\int_E \psi d\nu - \int_E \phi d\mu\right]$$

où le supremum est pris sur l'ensemble des fonctions (ψ,ϕ) vérifiant $\psi(x) \leq \phi(y) + d(x,y)$ pour tout $x, y \in E$. Même si la formulation est légèrement différente de celle du théorème 11.1.2, il s'agit bien du même ensemble de fonctions.

Ainsi, d'après l'inégalité de transport $T_1(C)$, nous avons

$$\int_E \psi d\nu - \int_E \phi d\mu \leq \sqrt{2C\mathrm{Ent}_\mu(\frac{d\nu}{d\mu})}.$$

Cette dernière inégalité peut s'écrire de manière équivalente sous la forme

$$\int_e \psi d\nu - \int_E \phi d\mu \leq \frac{C\lambda}{2} + \frac{1}{\lambda}\mathrm{Ent}_\mu(\frac{d\nu}{d\mu})$$

pour tout $\lambda > 0$. A présent, si $g = \frac{d\nu}{d\mu}$ cette nouvelle inégalité devient

$$\int_E fg d\mu \leq \mathrm{Ent}_\mu(g)$$

avec $f = \lambda\psi - \lambda\int_E \phi d\mu - \frac{C\lambda^2}{2}$. Puisque cette inégalité est satisfaite pour tout choix g (i.e. pour toute mesure de probabilité ν), il suffit de choisir $g = \frac{e^f}{\int_E e^f d\mu}$ pour obtenir que $\log\int_E e^f d\mu \leq 0$. Autrement dit,

$$\int_E e^{\lambda\psi} d\mu \leq e^{\lambda\int_E \phi d\mu + C\frac{\lambda^2}{2}}.$$

Si $F \in \mathrm{Lip}(1)$, il est alors possible de choisir $F = \psi = \phi$ dans ce qui précède (le couple (ψ,ϕ) ainsi défini vérifie bien la condition $\psi(x) \leq \phi(y) + d(x,y)$) afin de conclure. □

11.3.2 Inégalité de concentration indépendante de la dimension et $T_2(C)$

La démonstration précédente peut également être mise en place pour W_2 à l'aide du théorème de dualité 11.2.2.

Théorème 11.3.2 ($T_2(C)$ et concentration)**.** *Soit $\mu \in \mathcal{P}_2(\mathbb{R}^n)$. Alors*

$$W_2^2(\mu,\nu) \leq \sqrt{CH\nu|\mu)}$$

pour une constante $C > 0$ et pour tout $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ si et seulement si

$$\int_{\mathbb{R}^n} e^{Qf} d\mu \leq e^{\int_{\mathbb{R}^n} f d\mu}$$

pour toute fonction mesurable bornée $f : \mathbb{R}^n \to \mathbb{R}$; ci-dessus, Qf désigne l'infimum-convolution

$$Qf(x) = \inf_{y\in\mathbb{R}^n} \big[f(y) + \frac{1}{2C}|x-y|^2\big], \quad \textit{avec} \quad x \in \mathbb{R}^n$$

entre f et le coût quadratique.

Remarque. Lorsque F est lipschitzienne, il est possible de montrer que $QF \geq F - \frac{C}{2}\|F\|^2_{\mathrm{Lip}}$. Ainsi le théorème précédente fournit un résultat plus fort que le théorème 11.3.1. En fait, il a été démontré par Gozlan (cf. [66]) que l'inégalité $T_2(C)$ est équivalente à une inégalité de concentration gaussienne avec des paramètres indépendants de la dimension.

Comme mentionné dans la remarque précédente, l'inégalité de transport $T_2(C)$ se tensorise indépendamment de la dimension.

Proposition 11.3.1. *Soit $\mu = \mu_1 \otimes \dots, \otimes \mu_n$ une mesure de probabilité sur les boréliens de $\mathbb{R}^n$. Supposons que pour tout $i = 1, \dots, n$, μ_i satisfait une inégalité de transport quadratique*

$$W_2(\mu_i, \nu_i) \leq \sqrt{C_i H(\mu_i|\nu_i)},$$

pour n'importe quelle mesure ν_i sur $\mathbb{R}$ et $C_i > 0$. Alors,

$$W_2(\mu, \nu) \leq \sqrt{\max_{i=1,\dots,n} C_i H(\mu|\nu)},$$

pour n'importe quelle mesure de probabilité ν sur $\mathbb{R}^n$.

11.4 Transport optimal et inégalités fonctionnelles

Dans cette section, nous allons voir de quelle manière certaines inégalités fonctionnelles rencontrées dans ce cours peuvent être obtenues par des arguments de transport optimal. A titre d'exemple, débutons par l'inégalité $T_2(1)$ satisfaite par la mesure gaussienne.

Proposition 11.4.1 (Talagrand)**.** *La mesure gaussienne standard sur $\mathbb{R}^n$ γ_n vérifie une inégalité de transport $T(2)$ avec $C = 1$.*

Démonstration. D'après la propriété de tensorisation, il suffit de démontrer le cas $n = 1$. Soit $f \geq 0$ telle que $\int_{\mathbb{R}} f d\gamma_1 = 1$ et posons $d\nu = f d\mu$. Par simplicité, supposons que $f > 0$ sur tout $\mathbb{R}$. Ensuite, définissons l'application de transport monotone $T : \mathbb{R} \to \mathbb{R}$ par

$$\nu\big(]-\infty, T(x)]\big) = \gamma_1\big(]-\infty, x]\big), \quad x \in \mathbb{R},$$

de telle sorte à ce que ν soit la mesure image de γ par l'application T. La formule de changement de variable nous fournit l'équation suivante, dite de Monge-Ampère 11.1.2,

$$f\big(T(x)\big)T'(x)e^{-T(x)^2/2} = e^{-x^2/2} \quad \text{pour tout} \quad x \in \mathbb{R}.$$

Ainsi, en prenant le logarithme de l'équation précédente, nous obtenons, pour tout $x \in \mathbb{R}$,

$$\log f\big(T(x)\big) + \log T'(x) - \frac{1}{2}T(x)^2 = -\frac{1}{2}x^2$$

En intégrant cette égalité par rapport à la mesure γ et en utilisant le fait que $\nu = T_{\#}\gamma_1$, nous trouvons

$$\int_{\mathbb{R}} \log f d\nu = \frac{1}{2}\int_{\mathbb{R}} [T(x)^2 - x^2] d\gamma_1 - \int_{\mathbb{R}} \log T' d\gamma_1.$$

En outre, une intégration par partie fournit que

$$\int_{\mathbb{R}} x(T-x) d\gamma_1 = \int_{\mathbb{R}} (T'-1) d\gamma_1$$

d'où

$$\begin{aligned} H(\gamma_1, \nu) = \int_{\mathbb{R}} \log f d\nu &= \frac{1}{2}\int_{\mathbb{R}} |x - T(x)|^2 d\gamma_1 + \int_{\mathbb{R}} [T' - 1 - \log T'] d\gamma_1 \\ &\geq \frac{1}{2}\int_{\mathbb{R}} |x - T(x)|^2 d\gamma_1 \end{aligned}$$

puisque $y \mapsto y - 1 - \log y \geq 0$ lorsque $y \geq 0$. De plus, comme ν est l'image de γ_1 par l'application de transport T, la mesure image π de γ_1 par l'application $x \mapsto \big(x, T(x)\big)$ a pour marginales γ_1 et ν respectivement. Ceci entraine, par définition de W_2,

$$\frac{1}{2}\int_{\mathbb{R}} |x - T(x)|^2 d\gamma_1 = \iint_{\mathbb{R}^2} \frac{1}{2}|x-y|^2 d\pi \geq W_2^2(\gamma_1, \nu),$$

ce qui conclut la démonstration.

□

En fait, il se trouve que le cas n-dimensionel peut également être obtenu de manière similaire (sans utiliser de tensorisation). L'idée est d'utiliser l'application de transport monotone du théorème de Brenier 11.1.3.

Soient μ et ν deux mesures de probabilité sur $\mathbb{R}^n$, rappelons qu'une application $T : \mathbb{R}^n \to \mathbb{R}^n$ envoit μ sur ν (ou transporte μ sur ν) si ν est l'image de la mesure ν par T. Autrement dit, pour toute fonction borélienne positive bornée $f : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}} f(y) d\mu(y) = \int_{\mathbb{R}} f\big(T(x)\big) d\nu(x).$$

Si μ et ν admettent un moment d'ordre deux, une application T poussant μ sur ν est dite optimale par rapport à la distance de Kantorovich-Wasserstein W_2 si

$$W_2(\mu, \nu)^2 = \frac{1}{2}\int_{\mathbb{R}} |x - T(x)|^2 d\mu(x).$$

Un résultat fondamental de Brenier [35] et Mc Cann [107] (présenté plus tôt) assure que lorsque μ est absolument continu par rapport à la mesure de Lebesgue, il existe une fonction convexe ϕ telle que $T = \nabla\phi$ transporte μ sur ν de manière optimale (au sens précédent).

Soit $\mu = \gamma_n$ la mesure gaussienne standard dans $\mathbb{R}^n$ et supposons que $d\nu = f d\gamma_n$ avec $f \geq 0$ et $\int f \mathrm{d}\gamma_\mathrm{n} = 1$. Lorsque celui-ci fait sens, la formule de changement de variables dans le transport de γ_n à ν fournit l'équation de Monge-Ampère suivante :

$$f\big(T(x)\big)\det\big(\mathrm{Hess}\phi(x)\big)e^{-|T(x)|^2/2} = e^{-|x|^2/2} \quad \text{pour tout} \quad x \in \mathbb{R}^n,$$

où Hess ϕ désigne la hessienne de ϕ (nous ignorons les problèmes de régularité et le fait que $T = \nabla\phi$ ne pourrait exister que presque partout). En reproduisant la démonstration du cas uni-dimensionel et en utilisant le fait que

$$\log \mathrm{Det}\,\big(\mathrm{Hess}\,\phi(x)\big) \leq \Delta\phi - n = \Delta\big(\phi - \frac{|x|^2}{2}\big), \tag{11.4.1}$$

nous obterons que $W_2(\gamma_n, \nu) \leq \sqrt{H(\nu|\gamma_n)}$. Cette argument s'étend aisement aux mesures de probabilités $d\mu = e^{-V}dx$ avec un potentiel strictement convexe V.

Théorème 11.4.1. *Soit $d\mu = e^{-V}dx$ où $\mathrm{Hess}V(x) \geq \rho Id$, $c > 0$ uniformément en $x \in \mathbb{R}^n$. Alors, pour toute mesure de probabilité ν sur $\mathbb{R}^n$,*

$$W_2(\nu, \mu) \leq \sqrt{\frac{1}{\rho}H(\nu|\mu)}.$$

11.4.1 Inégalité de Sobolev logarithmique

L'utilisation de l'équation de Monge-Ampère a permis de démontrer que γ_n satisfait une inégalité de transport $T_2(1)$. Nous allons voir que le même type d'argument fonctionne au niveau de l'inégalité de Sobolev logarithmique. Avant cela, il peut-être utile de d'observer quelques faits concernant l'entropie d'une fonction et l'entropie relative de deux mesures.

Si $\int_{\mathbb{R}^n} f d\mu = 1$, alors

$$\mathrm{Ent}_\mu(f) = \int_{\mathbb{R}^n} \log f d\nu = H(\nu|\mu)$$

avec $d\nu = f d\mu$. Rappelons que μ vérifie une inégalité de Sobolev logarithmique s'il existe $C > 0$ telle pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{Ent}_\mu(f^2) \leq 2C \int_{\mathbb{R}^n} |\nabla f|^2 d\mu$$

Si nous substituons f^2 par $f > 0$ telle que $\int_{\mathbb{R}^n} f d\mu = 1$, l'inégalité de Sobolev logarithmique précédente s'écrit de manière équivalente sous la forme suivante

$$H(\nu|\mu) \leq \frac{C}{2}I(\nu)$$

avec $I(\nu) = I_\mu(f) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} d\mu$ l'information de Fisher de $d\nu = f d\mu$. Dans cette formulation, il devient possible de démontrer la proposition 9.4.3 à l'aide des outils du transport optimal.

Démonstration. Soit $f > 0$ telle que $\int_{\mathbb{R}^n} f d\mu = 1$ et considérons $d\nu = f d\gamma_n$. Considérons l'application de transport T optimale donnée par le théorème de Brenier 11.1.3 :

$$i.e. \quad T_{\#\nu} = \mu \quad \text{avec} \quad T = \nabla\phi$$

où $\phi : \mathbb{R}^n \to \mathbb{R}$ est une fonction convexe. Pour éviter des arguments de régularisation, supposons que tous les éléments mis en jeu sont suffisamment régulier de sorte que l'équation de Monge-Ampère suivante ait du sens. Celle-ci s'écrit

$$f(x)e^{-|x|^2/2} = e^{-|T(x)|^2/2}\mathrm{Det}\big(\mathrm{Hess}\,\phi(x)\big), \quad \text{pour tout} \quad x \in \mathbb{R}^n.$$

En prenant le logarithme de l'égalité précédente et en utilisant (11.4.1), nous obtenons

$$\begin{aligned} \log f(x) - \frac{1}{2}|x|^2 &= -\frac{1}{2}|T(x)|^2 + \log \mathrm{Det}\big(\mathrm{Hess}\,\phi(x)\big) \\ &\leq -\frac{1}{2}|T(x)|^2 + \Delta\big(\phi - \frac{|x|^2}{2}\big). \end{aligned}$$

Ainsi, puisque $T = \nabla\phi$, nous en déduisons que

$$\begin{aligned} \log f(x) &\leq -\frac{1}{2}|\nabla\phi - x|^2 - x\cdot(\nabla\phi - x) + \Delta\big(\phi - \frac{|x|^2}{2}\big) \\ &= -\frac{1}{2}|\nabla\phi - x|^2 + L\big(\phi - \frac{|x|^2}{2}\big) \end{aligned}$$

avec $L = \Delta - x\cdot\nabla$ le générateur infinitésimal du semi-groupe d'Ornstein-Uhlenbeck. A présent, intégrons la dernière inégalité obtenue par rapport à $d\nu = f d\gamma_n$, ceci fourni

$$\int_{\mathbb{R}^n} f\log f d\gamma_n \leq -\frac{1}{2}\int_{\mathbb{R}^n} f|\nabla\phi - x|^2 d\gamma_n + \int_{\mathbb{R}^n} fL\big(\phi - \frac{|x|^2}{2}\big)d\gamma_n.$$

Il est alors possible d'utiliser la formule d'intégration par partie vérifiée par L et γ_n,

$$\int_{\mathbb{R}^n} fL\big(\phi - \frac{|x|^2}{2}d\gamma_n = -\int_{\mathbb{R}^n} \nabla f\cdot\nabla(\phi - x)d\gamma_n$$

donc

$$\begin{aligned} \int_{\mathbb{R}^n} f\log f d\gamma_n &\leq -\frac{1}{2}\int_{\mathbb{R}^n} f|\nabla\phi - x|^2 d\gamma_n - \int_{\mathbb{R}^n} \nabla f\cdot\nabla(\phi - x)d\gamma_n \\ &\leq \frac{1}{2}\int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f}d\gamma_n = \frac{1}{2}I(\nu) \end{aligned}$$

La dernière ligne s'obtient en utilisant l'inégalité,

$$-\frac{1}{2}|b|^2 - a\cdot b \leq \frac{1}{2}|a|^2 \quad \text{(valable pour tous vecteurs} \quad a, b \in \mathbb{R}^n)$$

en choisissant $a = \frac{|\nabla f|}{\sqrt{f}}$ et $b = \sqrt{f}\nabla(\phi - x)$. □

Remarque. Bien entendu, la démonstration précédente fonctionne aussi pour des mesures log-concaves :

$$d\mu = e^{-V}dx$$

lorsque $\mathrm{Hess}V(x) \geq \rho I_d$ avec $\rho > 0$. En fait, pour de telles mesures μ, il a été démontré par Caffarelli (cf. [38]) que l'application de transport T envoyant γ_n vers μ (i.e. $T_{\#\gamma_n} = \mu$) est lipschitzienne avec $\|T\|_{lip} \leq \rho^{-\frac{1}{2}}$.

11.4.2 Inégalités de Brunn-Minkowski et de Prekopa-Leidler

Pour illustrer la polyvalence des méthodes de transport optimal, nous allons montrer qu'elles permettent d'obtenir l'inégalité de Brunn-Minkowski dont nous rappelons l'énoncé ci-dessous.

Théorème 11.4.2 (Brunn-Minkowski). *Pour tout ensembles boréliens bornés $A, B \in \mathbb{R}^n$, nous avons*

$$\mathrm{Vol}_n(A+B)^{1/n} \geq \mathrm{Vol}_n(A)^{1/n} + \mathrm{Vol}_n(B)^{1/n} \tag{11.4.2}$$

où $A = B = \{x + y \quad ; \quad x \in A,\, y \in B\}$ désigne la somme de Minkowski des ensembles A et B ; Vol_n désigne l'élément de volume de $\mathbb{R}^n$ (i.e. il s'agit de la mesure de Lebesgue).

L'équation (11.4.2) peut s'exprimer de manière équivalente sous la forme multiplicative suivante

$$\mathrm{Vol}_n(\theta A + (1-\theta)B) \geq \mathrm{Vol}_n(A)^{\theta}\mathrm{Vol}_n(B)^{1-\theta} \quad \forall \theta \in [0,1] \tag{11.4.3}$$

Remarque. Comme nous l'avons déjà mentionné plus tôt (cf. théorème 6.2.1), l'inégalité de Brunn-Minkowski permet de résoudre facilement le problème isopérimétrique euclidien. En effet, si B désigne une boule euclidienne de rayon $r > 0$ centrée en 0 alors l'inégalité (11.4.2) nous assure que

$$\mathrm{Vol}_n(A_r)^{1/n} = \mathrm{Vol}_n(A+B)^{1/n} \geq \mathrm{Vol}_n(A)^{1/n} + v(r)^{1/n}$$

où $v(r)$ désigne le volume d'une boule euclidienne de rayon $r > 0$. De plus, si D est une boule euclidienne de rayon $s > 0$ telle que $\mathrm{Vol}_n(A) = \mathrm{Vol}_n(D)$ alors, puisque $v^{1/n}$ est une application linéaire, nous avons

$$\mathrm{Vol}_n(A)^{1/n} + v(r)^{1/n} = v(s)^{1/n} + v(r)^{1/n} = v(s+r)^{1/n} = \mathrm{Vol}_n(D_r)^{1/n}.$$

Autrement dit, nous avons montré

$$\mathrm{Vol}_n(A_r) \geq \mathrm{Vol}_n(D_r) \quad \forall r > 0$$

ce qui correspond bien à la solution du problème isopérimétrique euclidien.

Le passage de la forme additive à la forme multiplicative n'est pas difficile, voici les arguments principaux :

1. Pour tout $\theta \in [0,1]$, il suffit d'appliquer (11.4.2) aux ensembles θA et $(1-\theta)B$ pour obtenir (11.4.3).
2. Réciproquement, il suffit d'appliquer (11.4.3) aux ensembles

   $$A' = \mathrm{Vol}_n(A)^{-1/n}A \quad \text{et} \quad B' = \mathrm{Vol}_n(B)^{-1/n}B$$

   puis de choisir

   $$\theta = \frac{\mathrm{Vol}_n(A)^{1/n}}{\mathrm{Vol}_n(A^{1/n} + \mathrm{Vol}_n(B)^{1/n}} \in [0,1]$$

   pour obtenir, par homogénéité, (11.4.2).

L'intérêt de cette formulation multiplicative est que celle-ci admet une version fonctionnelle plus générale. Il s'agit de l'inégalité de Prékopa-Leidler.

Théorème 11.4.3 (Prékopa-Leidler)**.** *Soit $\theta \in [0,1]$ et considérons u, v et w des fonctions positives et mesurables sur $\mathbb{R}^n$ telles que*

$$w\big(\theta x + (1-\theta)y\big) \geq u(x)^{\theta} v(y)^{1-\theta} \quad \textit{pour tous} \quad x, y \in \mathbb{R}^n \tag{11.4.4}$$

alors

$$\int_{\mathbb{R}^n} w dx \geq \Big(\int_{\mathbb{R}^n} u dx\Big)^{\theta} \Big(\int_{\mathbb{R}^n} v dx\Big)^{1-\theta} \tag{11.4.5}$$

Remarque. En particulier, le résultat précédent appliqué aux fonctions caractéristiques d'ensembles mesurables bornés A et B de $\mathbb{R}^n$ permet de retrouver la forme multiplicative (11.4.3) de Brunn-Minkowski. L'inégalité de Prekopa-Leidler peut-être étendue à un contexte Riemannien en utilisant la notion de géodésique et de courbure (cf. [43].

Démonstration. La démonstration proposée ci-dessous est due à F. Barthe (cf. [18]). Celle-ci repose sur une récurrence sur la dimension n.

Supposons que $n = 1$. Par homégénéité, nous pouvons supposer que $\int_{\mathbb{R}} u dx = \int_{\mathbb{R}} v dx = 1$ et, par des arguments de régularisation, nous pouvons supposer que u et v sont des fonctions continues strictement positives. Notons alors $d\mu = u dx$ et $d\nu = v dx$ et considérons l'application de transport $T : \mathbb{R} \to \mathbb{R}$ telle que

$$\nu\big(]-\infty, T(x)]\big) = \mu\big(]-\infty, x]\big) \quad \text{pour tout} \quad x \in \mathbb{R}.$$

Observons que T est croissante et différentiable, de plus $T_{\#\mu} = \nu$. L'équation de Monge-Ampère associée (qui n'est rien d'autre qu'un changement de variable) s'écrit

$$v\big(T(x)\big)T'(x) = u(x) \quad \text{pour tout} \quad x \in \mathbb{R}.$$

Posons alors $z(x) = \theta x + (1-\theta)T(x)$, nous avons alors $z'(x) = \theta + (1-\theta)T'(x)$. De plus, d'après l'inégalité arithmético-géométrique, nous avons

$$z'(x) \geq \big(T(x)\big)^{1-\theta} \quad \text{pour tout} \quad x \in \mathbb{R}. \tag{11.4.6}$$

Alors, d'après l'hypothèse (11.4.4) et (11.4.6), nous avons

$$\begin{aligned} \int_{\mathbb{R}} w dx = \int_{\mathbb{R}} x\big(z(x)\big)z'(x)dx \quad &\geq \quad \int_{\mathbb{R}} u(x)^{\theta} v\big(T(x)\big)^{1-\theta} z'(x)dx \\ &\geq \quad \int_{\mathbb{R}} u(x)^{\theta} v\big(T(x)\big)^{1-\theta} \big(T'(x)\big)^{1-\theta} dx \\ &= \quad \int_{\mathbb{R}} u dx = 1 \end{aligned}$$

puisque, par construction, $u(x)^{1-\theta} = \Big(v\big(T(x)\big)\big(T'(x)\big)\Big)^{1-\theta}$. Ceci démontre le résultat lorsque $n = 1$. Supposons alors que $n > 1$ et que le théorème de Brunn-Minkowski est vérifié sur $\mathbb{R}^{n-1}$. Considérons u, v et w trois fonctions vérifiant (11.4.4) pour un certain $\theta \in [0,1]$. Soit $q \in \mathbb{R}$ fixé et définissons $u_q : \mathbb{R}^{n-1} \to [0, +\infty[$ par $u_q(x) = u(x, q)$. Procédons de manière analogue pour définir v_q et w_q à partir de v et w. Observons ensuite que si $q = \theta q_0 + (1-\theta)q_1$ avec $q_0, q_1 \in \mathbb{R}$ alors

$$w_q\big(\theta x + (1-\theta)y\big) \geq u_{q_0}(x)^\theta v_{q_1}(y)^{1-\theta} \quad \text{pour tous} \quad x, y \in \mathbb{R}^{n-1}.$$

D'après l'hypothèse de récurrence, nous en déduisons que

$$\int_{\mathbb{R}^{n-1}} w_q dx \geq \Big(\int_{\mathbb{R}^{n-1}} u_{q_0} dx \Big)^\theta \Big(\int_{\mathbb{R}^{n-1}} v_{q_1} dx \Big)^{1-\theta}$$

ce qui permet d'appliquer les arguments mis en place en dimension 1. Autrement dit,

$$\int_{\mathbb{R}^n} w dx = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}^{n-1}} w_q dx) dq \geq \Big(\int_{\mathbb{R}^n} u dx)^\theta \Big(\int_{\mathbb{R}^n} vd \Big)^{1-\theta}$$

ce qui est le résultat voulu. □

Nous expliquons ci-dessous, de quelle manière l'inégalité de Prekopa-Leindler (vue comme version fonctionnelle de l'inégalité de Brunn-Minkowski) permet d'obtenir des inégalités de concentrations pour certains types de mesures (mesures uniformes sur la sphère ou mesure gaussiennes par exemple). Les idées présentées ci-dessous ont été introduites par Maurey (cf. [106]).

Supposons que μ est une mesure de probabilité sur $\mathbb{R}^n$ de la forme $d\mu = e^{-V} dx$ avec V un potentiel régulier strictement convexe. Autrement dit, il existe $\rho > 0$ telle que pour tout $x, y \in \mathbb{R}^n$,

$$V(x) + v(y) - 2V\big(\frac{x+y}{2}\big) \geq \frac{\rho}{4}|x-y|^2 \tag{11.4.7}$$

Un exemple typique de telle mesure est la mesure gaussienne γ_n sur $\mathbb{R}^n$ pour laquelle l'inégalité précédente est vérifiée pour $\rho = 1$. L'inégalité de Prékopa-Leidler nous permet de montrer le théorème suivant :

Théorème 11.4.4. *Soit μ une mesure de probabilité telle que $d\mu = e^{-V} dx$ et V vérifie (11.4.7). Alors,*

$$\alpha_\mu(r) \leq 2e^{-\rho r^2/4} \quad r > 0$$

avec α_μ la fonction de concentration associée à μ (cf. définition 6.1.1). En particulier,

$$\alpha_{\gamma_n}(r) \leq 2e^{-r^2/4} \quad r > 0$$

Démonstration. Considérons les fonctions suivantes :

$$u(x) = e^{Qf(x)-V(x)}, \quad v(y) = e^{-f(y)-V(y)} \quad \text{et} \quad w(z) = e^{-V(z)}$$

où, rappelons-le, Qf est l'infimum-convolution défini par

$$Qf(x) = \inf_{y \in \mathbb{R}^n} \big[f(y) + \frac{\rho}{4}|x-y|^2\big] \quad x \in \mathbb{R}^n.$$

Par définition de Qf et l'hypothèse de convexité (11.4.7), la condition (11.4.4) est satisfaite pour le choix $\theta = \frac{1}{2}$. En conséquence, d'après le théorème de Prékopa-Leidler 11.4.3, nous avons

$$1 = \int_{\mathbb{R}^n} e^{-V} dx \geq \int_{\mathbb{R}^n} e^{Qf} d\mu \int_{\mathbb{R}^n} e^{-f} d\mu.$$

Pour obtenir de la concentration, il suffit de choisir un ensemble mesurable $A \subset \mathbb{R}^n$ et d'appliquer ce qui précède à la fonction f définie par

$$f(x) = \begin{cases} 0 & \text{si} \quad x \in A \\ +\infty & \text{sinon.} \end{cases}$$

Dans ce cas, $Qf(x) = \frac{\rho}{4} d(x, A)^2$ avec $d(x, A)$ la distance euclidienne du point x à l'ensemble A. Tous ceci entraine alors que

$$\int_{\mathbb{R}^n} e^{\frac{\rho}{4} d(\cdot, A)^2} d\mu \leq \frac{1}{\mu(A)}$$

puisque $\int_{\mathbb{R}^n} e^{-f} d\mu = \mu(A)$. Enfin, pour tout $r > 0$, posons $F(x) = \min\big(d(x, A), r\big)$. Nous avons alors

$$1 - \mu(A_r) = \mu(F \geq r) \leq e^{-\lambda r^2} \mathbb{E}[e^{\lambda F^2}]$$

et le choix de $\lambda = \frac{\rho}{4}$ permet d'obtenir le résultat escompté

$$1 - \mu(A_r) \leq \frac{1}{\mu(A)} e^{-\rho r^2/4}.$$

□

11.5 Semi-groupe d'Hamilton-Jacobi et hypercontractivité

Certains résultats décrit plus tôt faisaient intervenir des fonctions obtenues par infimum-convolution (les fonctions Q_f). Il est possible d'expliquer plus en détails pourquoi de telles fonctions interviennent. Pour cela, nous devons introduire les équations aux dérivées partielles d'Hamilton-Jacobi.

Dans un contexte euclidien, ce problème s'énonce comme suit : étant donnée une fonction lipschitzienne $f : \mathbb{R}^n \to \mathbb{R}$, nous souhaitons trouver $(x, t) \mapsto v(x, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ telle que

$$\begin{cases} \partial_t v + \frac{1}{2} |\nabla v|^2 = 0 & \text{sur} \quad \mathbb{R}^n \times]0, +\infty[\\ v = f \quad \text{sur} \quad \mathbb{R}^n \times \{t = 0\} \end{cases}$$

La solution de ce problème est donnée par l'infimum-convolution de f avec la distance euclidienne :

$$Q_t f(x) = \inf_{y \in \mathbb{R}^n} \left[f(y) + \frac{1}{2t} |x - y|^2 \right] \quad \text{avec} \quad t > 0 \quad \text{et} \quad x \in \mathbb{R}^n.$$

Il se trouve que la famille d'opérateur $(Q_t)_{t \geq 0}$ défini un semi-groupe de générateur infinitésimal (non linéaire) $L = -\frac{1}{2} |\nabla f|^2$. Cette famille permet d'approfondir les liens entre transport optimal et inégalités fonctionnelles.

Débutons par un résultat de Otto-Villani qui permet d'obtenir une inégalité $T_2(C)$ à partir d'une inégalité de Sobolev logarithmique.

Théorème 11.5.1 (Otto-Villani)**.** *Considérons la mesure de probabilité $d\mu = e^{-V}dx$ avec $V : \mathbb{R}^n \to \mathbb{R}$. Si μ vérifie une inégalité de Sobolev logarithmique de constante $C > 0$:*

$$i.e. \quad \mathrm{Ent}_\mu(f^2) \leq 2C\mathbb{E}_\mu[|\nabla f|^2]$$

pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ suffisamment régulière, alors μ vérifie une inégalité de transport $T_2(C)$.

Démonstration. Pour démontrer ceci, nous utiliserons la formulation duale de $T_2(C)$ (cf. théorème 11.3.2) : μ vérifie une inégalité $T_2(C)$ de constante C si et seulement si

$$\int_{\mathbb{R}^n} e^{Qf} d\mu \leq e^{\int_{\mathbb{R}^n} f d\mu}$$

avec $Qf(x) = \inf_{y\in\mathbb{R}^n} \left[f(y) + \frac{1}{2C}|x-y|^2\right]$ pour tout $x \in \mathbb{R}^n$.. En outre, à l'aide des notations introduites plus tôt, observons que $Qf(x) = Q_C f(x)$; ceci expliquant pourquoi nous allons utiliser le semi-groupe d'Hamilton-Jacobi $(Q_t)_{t\geq 0}$ pour démontrer le théorème d'Otto-Villani.

A cet effet, observons que pour tout $t, s > 0$ nous avons $Q_t(sf) = sQ_{st}f$. En particulier, si $t = C$ et $s = \frac{1}{C}$ nous avons

$$\int_{\mathbb{R}^n} e^{Qf} d\mu \leq e^{\int_{\mathbb{R}^n} f d\mu} \qquad \Longleftrightarrow \qquad \int_{\mathbb{R}^n} e^{\frac{1}{C}Q_1(Cf)} d\mu \leq \int_{\mathbb{R}^n} e^{\int_{\mathbb{R}^n} f d\mu}$$

Quitte à remplacer f par $\frac{1}{C}f$, cela revient à

$$\int_{\mathbb{R}^n} e^{\frac{1}{C}Q_1(f)} d\mu \leq e^{\frac{1}{C}\int_{\mathbb{R}^n} f d\mu}$$

Considérons alors la fonction $H(t) = \frac{\log\Lambda(t)}{t}$ pour $t \in [0,1]$ avec $\Psi(t) = \int_{\mathbb{R}^n} e^{\frac{t}{C}Q_t(f)} d\mu$. En observant le fait

$$H(0) = \lim_{t\to 0} \frac{1}{t}\log\Psi(t) = \frac{1}{C}\int_{\mathbb{R}^n} f d\mu$$

nous constatons que l'inégalité $T_2(C)$ est équivalente à $H(0) \geq H(1)$. Ainsi, devons montrer que si une inégalité de Sobolev logarithmique est vérifiée alors la fonction $t \mapsto H(t)$ est décroissante sur $[0,1]$.

Ici, nous avons

$$H'(t) = -\frac{1}{t^2}\log\Psi(t) + \frac{1}{t}\frac{\Psi'(t)}{\Psi(t)}$$

C'est pourquoi

$$H'(t) \leq 0 \qquad \Longleftrightarrow \qquad t\Psi'(t) \leq \Psi(t)\log\Psi(t)$$

Or,

$$\Psi'(t)(t) = \int_{\mathbb{R}^n} \frac{1}{C} Q_t(f) e^{\frac{t}{C}Q_t(f)} d\mu + \frac{t}{C} \int_{\mathbb{R}^n} \partial_t Q_t(f) e^{\frac{t}{C}Q_t(f)} d\mu.$$

Alors, en utilisant le fait que $Q_t(f)$ vérifie l'équation d'Hamilton-Jacobi $\partial_t Q_t(f) + \frac{1}{2}|\nabla Q_t(f)^2 = 0$, nous obtenons

$$t\Psi'(t) = \int_{\mathbb{R}^n} \frac{t}{C} Q_t(f) e^{\frac{t}{C}Q_t(f)} d\mu - \frac{t^2}{2C} \int_{\mathbb{R}^n} |\nabla Q_t(f)|^2 e^{\frac{t}{C}Q_t(f)} d\mu.$$

Nous allons alors utiliser l'inégalité de Sobolev logarithmique pour majorer le deuxième terme du membre de droite de l'inégalité précédente. En effet, si $g^2 = e^{\frac{t}{C}Q_t(f)}$ celle-ci nous assure que

$$\begin{aligned} \mathrm{Ent}_\mu(g^2) &= \int_{\mathbb{R}^n} \frac{t}{C} Q_t(f) e^{\frac{t}{C}Q_t(f)} d\mu - \Psi(t) \log \Psi(t) \\ &\leq \frac{t^2}{2C} \int_{\mathbb{R}^n} |\nabla Q_t(t)|^2 e^{\frac{t}{C}Q_t(f)} d\mu \end{aligned}$$

Ce qui est équivalent à

$$\int_{\mathbb{R}^n} \frac{t}{C} Q_t(f) e^{\frac{t}{C}Q_t(f)} d\mu - \frac{t^2}{2C} \int_{\mathbb{R}^n} |\nabla Q_t(f)|^2 e^{\frac{t}{C}Q_t(f)} d\mu \leq \Psi(t) \log \Psi(t).$$

Autrement dit $t\Psi'(t) \leq \Psi(t) \log \Psi(t)$ qui est l'inégalité désirée. □

Nous avions déjà vu, dans le cas gaussien (cf. théorème 9.4.1), que le semi groupe d'Ornstein-Uhlenbeck était hypercontractif si et seulement si γ_n vérifie une inégalité de Sobolev logarithmique. Il se trouve qu'il est possible d'obtenir un lien similaire entre une mesure de probabilité μ et le semi groupe d'Hamilton-Jacobi.

Théorème 11.5.2 (Bakry-Gentil-Ledoux)**.** *Soit μ une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue. Si μ vérifie une inégalité de Sobolev logarithmique de constante $C > 0$ alors, pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ mesurable et bornée nous avons*

$$\left\| e^{Q_t f} \right\|_{a+t/C} \leq \| e^f \|_a \quad \textit{pour tout} \quad t \geq 0 \quad \textit{et} \quad a \in \mathbb{R}. \tag{11.5.1}$$

Réciproquement, si l'inégalité (11.5.1) est vérifiée pour un certain $a \neq 0$, pour tout $t \geq 0$ et toute fonction mesurables bornées $f : \mathbb{R}^n \to \mathbb{R}$ alors μ vérifie une inégalité de Sobolev logarithmique de constante $C > 0$.

Démonstration. L'idée clé est de suivre la démonstration du théorème de Nelson 9.4.1 à l'échelle exponentielle. Autrement dit, considérons la fonction $H(t) = \frac{1}{q(t)} \log \Psi(t)$ avec $q(0) = a$ et $\Psi(t) = \int_{\mathbb{R}^n} e^{q(t)Q_t(f)} d\mu$. A nouveau, pour obtenir l'hypercontractivité de $(Q_t)_{t\geq 0}$ il suffit de montrer que $H(t) \leq H(0)$.

Ici, nous avons

$$H'(t) = \frac{q'(t)}{q(t)} \log \Psi(t) + \frac{1}{q(t)} \frac{\Psi'(t)}{\Psi(t)}.$$

En outre,

$$\begin{aligned}\Psi'(t) &= \int_{\mathbb{R}^n} \left[q'(t)Q_t(f) + q(t)\partial_t Q_t(t)\right] e^{q(t)Q_t(f} dmu \\ &= q'(t)\int_{\mathbb{R}^n} Q_t(f)e^{q(t)Q_t(f)}d\mu - \frac{q(t)}{2}\int_{\mathbb{R}^n} |\nabla Q_t f|^2 e^{q(t)Q_t(f)}d\mu\end{aligned}$$

puisque $Q_t(f)$ vérifie l'équation d'Hamilton-Jacobi $\partial_t Q_t(f) + \frac{1}{2}|\nabla Q_t(f)^2 = 0$. De plus,

$$\int_{\mathbb{R}^n} Q_t(f)e^{q(t)Q_t(f)} = \frac{1}{q(t)}\left[\mathrm{Ent}_\mu(g^2) + \Psi(t)\log\Psi(t)\right]$$

avec $g^2 = e^{q(t)Q_t(f)}$. En d'autres termes nous avons,

$$H'(t) = \frac{q'(t)}{q^2(t)\Psi(t)}\mathrm{Ent}_\mu(g^2) - \frac{1}{2\Psi(t)}\int_{\mathbb{R}^n} |\nabla Q_t f|^2 e^{q(t)Q_t(f)}d\mu$$

Or, d'après l'inégalité de Sobolev logarithmique appliquée à g^2, nous avons également

$$\mathrm{Ent}_\mu(g^2) \leq \frac{Cq^2(t)}{2}\int_{\mathbb{R}^n} |\nabla Q_t f|^2 e^{q(t)Q_t(f)}d\mu.$$

En conséquence,

$$H'(t) \leq 0 \quad\Longleftrightarrow\quad \frac{C}{2}q(t') - \frac{1}{2} \leq 0$$

La résolution de cette inégalité différentielle nous fournit $q(t) \leq \frac{t}{C} + a$ qui est le résultat attendu. La réciproque s'obtient de la même manière que dans le théorème 9.4.1. □

11.6 Hiérarchie des inégalités fonctionnelles

Comme nous l'avons déjà observé, il existe une certaine hiérarchie parmi les inégalités fonctionnelles étudiées. Au sommet de celle-ci, il y a l'inégalité de Sobolev logarithmique qui permet d'obtenir une inégalité de Poincaré. De plus, puisqu'une inégalité de Sobolev logarithmique entraine à la fois une inégalité de transport $T_2(C)$, il est naturel de s'intéresser aux liens existants entre l'inégalité $T_2(C)$ et l'inégalité de Poincaré. Le résultat suivant fournit une réponse à cette question

Proposition 11.6.1. *Soit μ une mesure de probabilité admettant une densité par rapport à la mesure de Lebesgue. Si μ satisfait une inégalité de transport $T_2(C)$ avec $C > 0$ alors μ vérifie une inégalité de Poincaré de constante $\frac{C}{2}$.*

Démonstration. Supposons qu'une inégalité de transport $T_2(C)$ est vérifiée. Dans sa formulation duale, cela signifie que

$$\int_{\mathbb{R}^n} e^{Q(f)}d\mu \leq e^{\int_{\mathbb{R}^n} d\mu}$$

Appliquons cette inégalité à tf avec $t \to 0$ et utilisons le fait suivant (valable pour tout $x \in \mathbb{R}^n$)

$$Q(tf)(x) = tf(x) - \frac{C}{4}t^2|\nabla f|^2 + o(t^2) \quad \text{lorsque} \quad t \to 0$$

afin de déterminer des développements limités des deux membres de la formulation duale de l'inégalité $T_2(C)$. D'une part, nous obtenons

$$\int_{\mathbb{R}^n} e^{Q(tf)}d\mu = 1 + t\int_{\mathbb{R}^n} fd\mu + \frac{t}{2}\int_{\mathbb{R}^n} f^2 d\mu - \frac{Ct^2}{4}\int_{\mathbb{R}^n} |\nabla f|^2 d\mu + o(t^2)$$

et, d'autre part,

$$e^{t\int_{\mathbb{R}^n} fd\mu} = 1 + t\int_{\mathbb{R}^n} fd\mu + \frac{t^2}{2}\Big(\int_{\mathbb{R}^n} fd\mu\Big)^2 + o(t^2).$$

En combinant ces deux résultats, et après simplifications, nous obtenons

$$\frac{t^2}{2}\mathrm{Var}_\mu(f) \leq \frac{Ct^2}{4}\int_{\mathbb{R}^n} |\nabla f|^2 d\mu$$

Autrement dit, une inégalité de Poincaré de constante $\frac{C}{2}$ est satisfaite. □

Remarque. D'une certaine manière, l'inégalité $T_2(C)$ peut se voir comme une version duale de la propriété (τ) introduite par Maurey dans [106].

Dans certains cas, il est possible de renverser certaines parties de la hiérarchie que nous venons d'établir afin d'obtenir une inégalité de Sobolev logarithmique à partir d'une inégalité de transport $T_2(C)$. Dans ce qui suit, nous considérons une mesure sur les boréliens de $\mathbb{R}^n$ de la forme $d\mu = e^V dx$ avec $V : \mathbb{R}^n \to \mathbb{R}$ une application convexe.

Théorème 11.6.1 (HWI (Otto-Villani))**.** *Dans le cadre précédent, supposons que μ vérifie une inégalité de transport $T_2(C)$ avec $C > 0$ alors l'inégalité suivante est satisfaite : pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ suffisamment régulière et pour toutes mesures $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ nous avons*

$$\mathrm{Ent}_\mu(f) \leq \sqrt{CT_2(\mu|\nu)}I^{1/2}(\nu) \tag{11.6.1}$$

où $I(\nu)$ est l'information de Fisher associée à la mesure ν.

Remarque.

1. En particulier, cette inégalité reliant entropie, information de Fisher et distance de Kantorovich-Wasserstein permet de renverser la hiérarchie habituelle des inégalités fonctionnelles. En effet, si une inégalité de transport $T_2(C)$ est satisfaite, nous obtenons (à partir de (11.6.1))

$$\mathrm{Ent}_\mu(f) \leq 2CI(\nu)$$

Autrement dit, μ vérifie une inégalité de Sobolev logarithmique de constante $2C$.

2. En fait, il est possible d'obtenir une inégalité HWI plus générale. En remplaçant l'hypothèse de convexité (portant sur V) par $\mathrm{Hess}\, V \geq \rho I_d$ avec $\rho \in \mathbb{R}$ (au lieu de $\rho \geq 0$), il est possible de montrer que

$$\mathrm{Ent}_\mu(f) \leq T_2(\mu,\nu)I^{1/2}(\nu) - \frac{\rho}{2}T_2^2(\mu,\nu)$$

De plus, si μ vérifie une inégalité de transport $T_2(C)$ avec une constante $C > 0$ suffisamment petite, il est possible de montrer que μ vérifie également une inégalité de Sobolev logarithmique ; cependant le lien entre les constantes devient plus complexe (cf. [14]).

Démonstration. D'après le théorème de Brenier, il existe une application convexe $\phi : \mathbb{R}^n \to \mathbb{R}^n$ telle que

$$T_2^2(\nu,\mu) = \int_{\mathbb{R}^n} |x - \nabla\phi(x)|^2 d\nu(x)$$

avec $d\nu = f d\mu$. L'équation de Monge-Ampère associé s'écrit alors

$$f(x)e^{-V(x)} = \big|\mathrm{Det}\big(\mathrm{Hess}\,\phi(x)\big)\big|e^{-V\big(\nabla\phi(x)\big)} \quad x \in \mathbb{R}^n$$

Pour simplifier les notations, posons $\Theta(x) = \phi(x) - \frac{1}{2}|x|^2$ pour tous $x \in \mathbb{R}^n$. L'équation de Monge-Ampère se formule alors de manière équivalente sous la forme suivante :

$$f(x)e^{-V(x)} = \big|\mathrm{Det}\big(\mathrm{Id} + \mathrm{Hess}\,\Theta(x)\big|e^{-V\big(x+\nabla\Theta(x)\big)} \quad x \in \mathbb{R}^n.$$

Ainsi,

$$\begin{aligned} \log f(x) &= V(x) - V\big(x + \nabla\Theta(x)\big) + \log\Big(\big|\mathrm{Det}\big(\mathrm{Id} + \mathrm{Hess}\,\Theta(x)\big)\big|\Big) \\ &\leq V(x) - V\big(x + \nabla\Theta(x)\big) + \Delta\Theta(x) \end{aligned}$$

puisque $\log\big|\mathrm{Det}(I_d + A)\big| \leq \mathrm{Tr}(A)$ pour toutes matrices A symétriques réelles. En outre, puisque $x \mapsto V(x)$ est convexe,

$$V(x) - V\big(x + \nabla\Theta(x)\big) \leq -\nabla V(x)\cdot\nabla\Theta(x)$$

Si $L = \Delta - \nabla V \cdot V$ nous avons donc montré que

$$\log f(x) \leq L\Theta(x). \tag{11.6.2}$$

L'interêt de cette nouvelle formulation est que $d\mu = e^{-V(x)}dx$ est la mesure réversible et invariante associée à l'opérateur L, en particulier : la formule d'intégration par partie suivante est satisfaite $\int_{\mathbb{R}^n} f(-L\Theta)d\mu = \int_{\mathbb{R}^n} \nabla\Theta\cdot\nabla f d\mu$. Ainsi, en multipliant (11.6.2) par f et en intégrant par rappport à $d\mu$ nous obtenons

$$\int_{\mathbb{R}^n} f\log f d\mu \leq \int_{\mathbb{R}^n} fL\Theta d\mu = -\int_{\mathbb{R}^n} \nabla f\cdot\nabla\Theta d\mu$$

enfin, d'après l'inégalité de Cauchy-Schwarz, nous avons

$$\begin{aligned} -\int_{\mathbb{R}^n} \nabla f\cdot\nabla\Theta d\mu &\leq \sqrt{\int_{\mathbb{R}^n} |\nabla\Theta|^2 f d\mu} \times \sqrt{\int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f}} \\ &= T_2(\mu,\nu)I^{1/2}(\nu) \end{aligned}$$

puisque $T_2^2(\mu,\nu) = \int_{\mathbb{R}^n} |x - \nabla\phi(x)|^2 d\nu = \int_{\mathbb{R}^n} |\nabla\Theta|^2 f d\mu$. □

Remarque. Nous venons de proposer une démonstration utilisant des arguments provenant de la théorie du transport optimal. Il également possible d'obtenir une preuve qui n'utilise que des semi-groupes (cf. [14]). Voici les arguments clés :

1. Par interpolation, il est possible de montrer que pour tout $T > 0$

$$\mathrm{Ent}_\mu(f) = \int_0^T I_\mu\big(P_t(f)\big)dt + \mathrm{Ent}_\mu\big(P_T(f)\big)$$

   avec l'information de Fisher de la fonction f sous μ $I_\mu(f) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} d\mu$ et $(P_t)_{t\geq 0} = (e^{tL})_{t\geq 0}$, le semigroupe associé à l'opérateur $L = \Delta - \nabla V \cdot \nabla$ et la mesure $d\mu = e^{-V}dx$. Cette égalité peut également être vue comme une conséquence de l'identité de Bruijn :

$$\frac{d}{dt}\mathrm{Ent}_\mu(P_t f) = -I_\mu(P_t f) \tag{11.6.3}$$

2. Lorsque $\mathrm{Hess}(V) \geq \rho I_d$ avec $\rho \in \mathbb{R}$, cette hypothèse entraine une décroissante exponentielle de l'information de Fisher le long du semi groupe. Autrement dit,

$$I_\mu(P_t f) \leq e^{-2\rho t} I_\mu(f) \quad t \geq 0$$

   Ceci permet de contrôler le premier terme impliquant l'information de Fisher.
3. Enfin pour conclure, il faut contrôler $\mathrm{Ent}_\mu\big(P_T(f)\big)$ à l'aide d'inégalité d'Harnack. Celle-ci fait intervenir le semi groupe d'Hamilton-Jacobi $(Q_t)_{t\geq 0}$. Il est alors possible de conclure la preuve en utilisant la réversibilité de $(Q_t)_{t\geq 0}$ par rapport à μ avec la formulation duale de l'inégalité $T_2(C)$.

Détaillons ce dernier point. Il a été démontré (cf. [145, 146, 13]) qu'un critère $CD(\rho, +\infty)$ est équivalente à une inégalité de type Harnack. Dans sa forme logarithmique, celle-ci s'énonce comme suit

$$P_t(\log g)(x) \leq \log P_t g(y) + \frac{\rho}{2}\frac{|x-y|^2}{e^{2\rho t} - 1} \quad \text{pour tout} \quad x, y \in \mathbb{R}^n \quad \text{et} \quad t \geq 0.$$

Ce genre de résultat est notamment pratique pour obtenir des minorations, sous μ, du noyau $p_t(x,y)$ associé au semi-groupe. Appliquons l'inégalité précédente avec $t = T > 0$ et $g = P_T(f)$, nous obtenons

$$P_T(\log P_T f)(x) \leq \log P_{2T} f(y) + \frac{|x-y|^2}{2\beta(T)} \quad \text{pour tout} \quad x, y \in \mathbb{R}^n$$

où

$$\beta(T) = \begin{cases} \frac{e^{2\rho T}-1}{\rho} & \text{si} \quad \rho \neq 0 \\ 2T & \text{sinon.} \end{cases}$$

En prenant l'infimum sur $y \in \mathbb{R}^n$, ceci-fournit $P_T \log P_T(f)(x) \leq Q_{\beta(T)} \log P_{2T} f$ avec $(Q_t)_{t\geq 0}$ le semi-groupe d'Hamilton-Jacobi introduit plus tôt dans ce chapitre.

Par réversibilité, nous avons alors $\mathrm{Ent}_\mu(P_T f) = \int_{\mathbb{R}^n} P_T f \log P_T f d\mu = \int_{\mathbb{R}^n} f P_T \log P_T f d\mu$. C'est pourquoi, en utilisant l'inégalité d'Harnack logarithmique, nous obtenons

$$\mathrm{Ent}_\mu(P_T f) \leq \int_{\mathbb{R}^n} f Q_{\beta(T)}(\log P_{2T} f) d\mu = \frac{1}{\beta(T)} \int_{\mathbb{R}^n} f Q_1(\beta(T) \log P_{2T} f) d\mu.$$

En outre, $\int_{\mathbb{R}^n} \log P_{2T} f d\mu \leq \log(\int_{\mathbb{R}^n} P_{2T} f d\mu) = 0$. Autrement dit,

$$\mathrm{Ent}_\mu(P_T f) \leq \frac{1}{\beta(T)}\left[\int_{\mathbb{R}^n} Q_1 f d\nu - \int_{\mathbb{R}^n} f d\mu\right].$$

En prenant le supremum sur les fonctions f, nous en déduisons (via la formulation duale de W_2 11.2.2) que

$$\mathrm{Ent}_\mu(P_T f) \leq \frac{1}{2\beta(T)} W_2(\nu, \mu)^2$$

avec $d\nu = f d\mu$. En utilisant cette dernière majoration avec les deux premiers points, il ne reste plus qu'à optimiser en $T > 0$ pour conclure la démonstration.

11.7 Références historiques

Les notes qui suivent sont tirées de [143, 14, 68].

Le transport optimal trouve ses sources dans les travaux de Monge à la fin du 18ième siècle. Grâce à ses capacités mathématiques, ce brillant géomètre fut admit dans une école militaire alors que ses origines modestes auraient du l'exclure de ce genre d'instruction. Là-bas, il peut développer ses idées qui lui permirent de mettre en place de nouvelles méthodes qui s'avérèrent efficaces et polyvalentes. Tout ceci, lui permit d'être nommé professeur à l'âge de 22 ans ; cette promotion assurait également à ses supérieurs que les travaux de Monge deviennent des secrets militaires, uniquement accessibles à des officiers de haut rang. Par la suite, Monge devient l'un des ardents combattants scientifiques de la révolution française et servi, en tant que professeur, sous de nombreux régimes. Il échappa notamment à une sentence de mort durant la terreur et, plus tard, devint un ami proche de Napoléon. Il enseigna d'ailleurs à l'école normale supérieur et l'école polytechnique de Paris qui venaient d'être reformées au début du règne napoléonien.

En 1781, Monge publia son fameux mémoire sur la théorie des déblais et des remblais. Dans ce mémoire, il considère le problème suivant : supposons que nous disposions à un endroit E d'une quantité de sable que nous souhaitons déplacer sur un site de construction F. De plus, chacun des déplacements d'une partie du sable x de E à un emplacement y sur le site F coûte une certaine somme $c(x, y)$. En modélisant ceci (le tas de sable et le site de construction) par des mesures de probabilités μ et ν, la théorie du transport optimal cherche à minimiser le coût (par la suite ce coût sera la distance d entre les points $x \in E$ et $y \in F$) d'un tel déplacement.

D'un point de vue économique, le problème précédent peut aussi s'énoncer en termes de boulangeries (par exemple) cherchant à approvisionner en pain ou pâtisseries différents cafés. Dans ce problème, les quantités de pains produites dans ces boulangeries ainsi que les quantités consommées dans les cafés sont supposés connus à l'avance. Le problème revient donc à trouver de quelle manière chaque unité de pain doit être acheminée de sorte que le coût de ce transport soit minimal.

Beaucoup plus tard, le problème de Monge fut redécouvert par la mathématicien Kantorovich. Ce dernier, né en 1912, est également un brillant mathématicien : à l'âge de 18, il possédait déjà une solide renommée internationale et devint professeur à l'âge de 22 ans. En 1938, un laboratoire le consulta à propos de problèmes d'optimisation en économie. Kantorovich développa alors de nouvelles méthodes de programmation linéaire afin de les résoudre et ces nouveaux travaux firent de lui une personnalité incontournable de l'économie. Notons toutefois qu'une grande partie de ses travaux ne furent pas publiés immédiatement car ces derniers étaient étroitement surveillés par les autorités Soviétiques ; d'ailleurs, il n'était même pas permit à Kantorovich de s'exprimer publiquement sur ses découvertes. Enfin, En 1975, ses travaux furent mondialement connus et Kantorovich fut récompensé (conjointement avec Koopmans) pour leurs contributions à la théorie d'optimisation d'allocation de ressources.

Les travaux que nous avons présentés dans ce chapitre concernent le problème de couplages optimaux. Dans ce domaine, Kantorovich fut à l'origine de nombreux résultats qu'il démontra à l'aide d'outils d'analyse fonctionnelle. Citons en particulier le théorème de dualité qui porte son nom. Il fallut attendre encore quelques années pour que Kantorovich fasse le lien entre ses travaux et ceux de Monge dans [80, 81, 82]. Ces travaux peuvent être considérés comme l'acte de naissance de la théorie du transport optimal ; depuis, le problème de couplage optimal porte le nom de problème de Monge-Kantorovich. Cette théorie possède de nombreux liens avec d'autres domaines des mathématiques, citons quelques exemples : la reconnaissance de forme, la conception d'antennes réflectives ou de lentilles, problème d'irrigations et de conception de réseaux, ...(cf. [143]).

L'existence d'une application de transport provenant du gradient d'une fonction convexe prend ses origines dans les travaux de Knott et Smith (cf. [123]) mais aussi ceux de Rüschendorf et Rachev (cf. [119]). Brenier, quant à lui, annonça ses résultats majeurs dans une courte note pour ensuite publier les démonstrations de ces derniers dans [35] en 1991 et mis en avant la pertinence géométrique de son théorème dans l'étude d'équations aux dérivées partielles.

Le choix de la terminologie des distances de Wasserstein est, apparemment, attribué à Dobrushin. Cependant, il semblerait que la définition explicite de ces distances n'apparaisse pas exactement dans les travaux de Wasserstein et furent plutôt l'oeuvre de Kantorovich. Bien que cette terminologie de distance de Wasserstein soit extrêmement répandue actuellement, il semble judicieux de lui attribuer le nom de distance de Kantorovich-Wasserstein ; malgré tout, ceci ne permet pas de rendre hommage aux nombreux mathématiciens qui ont travaillé sur cette notion : Gini, Mallows, Tanaka, Hoeffding, Fréchet, Rubinstein, Ornstein, Salvemini, Dall'Algio,...

Parmi les distances de Wasserstein, le cas $p = 1$ a déjà été rencontré dans un chapitre précédent sous le nom de distance de Kantorovich-Rubinstein permettant de considérer l'espace des mesures signées $\mathcal{M}(E)$ comme un espace métrique.

Nous avons retrouver dans ce chapitre l'inégalité de Brunn-Minkowski sous sa forme fonctionnelle d'inégalité de Prékopa-Leindler. Alors que sa formulation euclidienne semblait bien connu des mathématiciens de l'époque (cf. [98, 114]), il fallut attendre les travaux de Cordero-Erausquin pour obtenir, grâce au formalisme du transport optimal, une version de celle-ci sur la sphère. Plus tard, en 2001, Cordero-Erausquin, McCann et Schmuckenschläger furent capable de développer les outils nécessaires permettant d'établir rigoureusement ce genre d'inégalité dans un cadre riemannien [43]. Bobkov et Ledoux ont montré dans [26] de quelle manière il était possible de retrouver de nombreuses inégalités fonctionnelles (notamment celle de Sobolev logarithmique) à partir d'une inégalité semblable à celle de Prékopa-Leindler.

Les inégalités de transport $T_p(C)$ furent étudiées dès le début des années 90 ; en revanche, l'inégalité de Csiszar-Pinsker date des années 60. La formulation duale de ces inégalités, pour $p = 1$ et $p = 2$ est due à Bobkov et Götze (cf. [23]). L'utilisation systématique de cette nouvelle formulation fut exploitée par Bobkov, Gentil et Ledoux dans [25]. Ce genre d'argument est notamment à la base de la démonstration du théorème d'Otto-Villani, assurant qu'une inégalité de Sobolev logarithmique entraine une inégalité de transport $T_2(C)$. Toujours dans [25], les auteurs mettent également en avant les liens existants entre inégalités de Sobolev logarithmiques et hypercontractivité du semi-groupe d'Hamilton-Jacobi. L'argument de tensorisation remonte aux travaux de Marton et ont été adaptés à de nouvelles situations (comme celle des chaines de Markov faiblement dépendantes) depuis. Le théorème de Gozlan, établissant l'équivalence entre une inégalité $T_2(C)$ avec de la concentration gaussienne indépendante de la dimension, introduit une idée nouvelle : utiliser le théorème de Sanov provenant de la théorie des grandes déviations ; des utilisations de ce genre d'idée sont décrites dans l'article de survol [68], citons, par exemple, la découverte par Gozlan d'une nouvelle démonstration du théorème d'Otto-Villani.

Comme nous l'avons mentionné dans ce chapitre, Talagrand fut le premier à montrer que la mesure gaussienne vérifiait une inégalité de transport quadratique. L'extension de ce type d'inégalités à des mesures strictement log-concaves a été entrepris ensuite, citons notamment les travaux de Gentil, Guillin et Miclo à ce sujet. L'inégalité HWI d'Otto et Villani permit d'identifier les liens profonds entre le transport optimal et le monde des inégalités fonctionnelles. A ce titre, la démonstration de l'inégalité de Sobolev logarithmique gaussienne et due à Cordero-Erausquin. Nous avons aussi vu dans ce chapitre qu'il était possible, sous certaines hypothèses, d'obtenir une inégalité de Sobolev logarithmique à partir d'une inégalité de transport $T_2(C)$ (via l'inégalité HWI), il est donc naturel de s'interroger quant à une éventuelle équivalence entre ces deux inégalités. Cattiaux et Guillin montrent dans [39], à l'aide d'un contre-exemple, que l'inégalité de Sobolev logarithmique est strictement plus forte qu'une inégalité de transport de coût quadratique.

Dans cette courte introduction à la théorie du transport optimal, nous n'avons pas abordé les travaux récents de Lott, Sturm et Villani qui ont utilisé le formalisme du transport optimal pour proposer une définition synthétique de courbure dans des espaces métriques mesurés (cf. [143]).