Transport optimal de mesures et quelques applications

Thomas CAVALLAZZI

École Normale Supérieure de Rennes

Sous la supervision de François BOLLEY

Laboratoire de Probabilités, Statistiques et Modélisation, Sorbonne Université

Introduction

Ce travail présente les bases de la théorie du transport optimal. On peut faire remonter la naissance de la théorie du transport optimal au mathématicien et homme politique français Gaspard Monge. Ce dernier se posait la question de réaliser le transport du sable issu de différentes carrières situées à divers endroits, et ceci de manière à minimiser le coût de transport global. Monge a formalisé le problème en 1781. Au vingtième siècle, le mathématicien et économiste soviétique Leonid Kantorovich a prouvé l'existence d'une solution optimale au problème général de transport, et a développé la théorie, ce qui lui valut le prix Nobel d'économie en 1975. La théorie du transport est toujours en plein développement grâce à ses applications dans de nombreux domaines. Des avancées importantes ont été faites par le mathématicien italien Alessio Figalli, qui a reçu la médaille Fields en août 2018. En plus d'avoir contribué à développer la théorie du transport, ce dernier l'a utilisée pour étudier les déformations de cristaux et les équations semi-géostrophiques utilisées en météorologie.

La première partie de ce travail présente le problème du transport optimal et sa résolution. La seconde partie a pour but de définir les distances de Wasserstein, qui sont des distances sur certains espaces de probabilités. Ces distances sont directement liées au problème de transport optimal. La fin de ce travail illustre l'utilité des outils précédemment développés appliqués à l'étude du phénomène de relaxation vers l'équilibre de l'équation de Boltzmann homogène en espace. Cette équation décrit l'évolution d'un gaz dilué spatialement homogène au cours du temps. On montrera qu'asymptotiquement, la répartition statistique des vitesses dans le gaz est gaussienne, quel que soit l'état initial du gaz.

Je tiens à remercier François Bolley pour le temps qu'il a consacré à encadrer mon stage, les réponses à mes (nombreuses) questions, et pour m'avoir fait découvrir de belles mathématiques.

Table des matières

1 Transport optimal de mesures

1.1 Les problèmes de Monge et de Kantorovich

1.1.1 Le problème de Monge

Le problème initial posé par Gaspard Monge en 1781 consiste au transport d'un tas de sable d'un certain volume sur un trou à remplir du même volume, disons 1. On peut représenter le sable et le trou par deux probabilités μ et ν.

Dans un premier temps, on suppose qu'il existe une application de transport T, c'est-à-dire que le sable initialement à la position x est envoyé entièrement en $T(x)$. Pour réaliser le transport du sable, il faut que le volume sous une partie du trou à l'arrivée, disons B, soit égal au volume du sable au départ qui sera transporté sur B par T. On veut donc imposer :

$$\nu(B) = \mu(T^{-1}(B)),$$

qu'on peut interpréter comme le fait que ν est la mesure image de μ par T. Le cadre rigoureux pour poser le problème est celui des espaces polonais, c'est-à-dire des espaces métriques séparables et complets.

Notation : Dans la suite, on notera $\mathcal{P}(X)$ l'ensemble des mesures de probabilité sur un espace polonais X.

On peut maintenant poser rigoureusement le problème de Monge.

Problème de Monge : Soient X et Y deux espaces polonais, $\mu \in \mathcal{P}(X)$ et $\nu \in \mathcal{P}(Y)$. On suppose que ν est la mesure image de μ par une application mesurable $T : X \to Y$, ce qu'on notera :

$$\nu = T * \mu.$$

T est appelée application de transport. Enfin, on se donne une fonction coût de transport : $c : X \times Y \to \mathbb{R}^+$ mesurable. Ainsi $c(x, y)$ représente le coût unitaire de transport d'une unité de volume de x à y. Typiquement, si $X = Y$, c peut-être la distance sur l'espace métrique X. Le coût global de transport de la mesure μ sur la mesure ν via T est défini par :

$$\int_X c(x, T(x))\, d\mu(x).$$

Le problème de Monge consiste alors à minimiser le coût global de transport par rapport à l'application de transport T. On cherche une application de transport T, si elle existe, qui vérifie :

$$\inf_{T*\mu=\nu} \left\{ \int_X c(x, T(x))\, d\mu(x) \right\} = \int_X c(x, T(x))\, d\mu(x) \quad \text{(M)}.$$

Remarque 1.1. *Dans le cas où μ et ν admettent des densités strictement positives f et g par rapport à la mesure de Lebesgue sur $\mathbb{R}$, alors il existe une application de transport T. Si on note F et G les fonctions de répartition de μ et ν, alors l'application $T = G^{-1} \circ F$ transporte μ sur ν. En effet, pour tout $y \in \mathbb{R}$:*

$$T * \mu(]-\infty, y]) = F(F^{-1}(G(y))) = G(y) = \nu(]-\infty, y])$$

Malheureusement il n'existe pas toujours d'application de transport entre deux probabilités. Typiquement, aucune application ne transporte la masse de Dirac δ_0 sur $\frac{\delta_0+\delta_1}{2}$. Il est donc nécessaire de trouver une formulation plus générale, pour laquelle le problème est bien posé.

1.1.2 Généralisation : le problème de Kantorovich

On va d'abord définir rigoureusement le problème de Kantorovich et nous allons voir en quoi il généralise le problème de Monge.

Notation : Soient X et Y des espaces polonais, $\mu \in \mathcal{P}(X)$ et $\nu \in \mathcal{P}(Y)$. On note $\Pi(\mu, \nu)$ l'ensemble des probabilités π sur $X \times Y$ ayant comme lois marginales μ et ν. De telles probabilités sont appelées plans de transport envoyant μ sur ν.

Problème de Kantorovich : On considère deux espaces polonais X et Y, ainsi que deux probabilités μ et ν sur X et Y respectivement. On se donne également une fonction coût de transport $c : X \times Y \to \mathbb{R}^+$ mesurable. Le problème de Kantorovich correspond à la recherche d'un $\pi \in \Pi(\mu, \nu)$ réalisant :

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c \, d\pi := \inf_{\pi \in \Pi(\mu,\nu)} I[\pi] \quad \text{(K)}.$$

Une telle probabilité π est appelée plan de transport optimal et la valeur de la borne inférieure est appelée coût de transport optimal. Cette fois, le problème de l'existence d'une application de transport rencontré pour le problème de Monge ne se pose pas puisque $\mu \otimes \nu \in \Pi(\mu, \nu)$, qui est donc non vide. Dans la suite, nous démontrerons l'existence d'un plan de transport optimal.

Voyons en quoi le problème de Kantorovich généralise celui de Monge. On suppose qu'il existe une application de transport $T : X \to Y$. On considère la mesure image de μ par l'application mesurable $Id_X \times T$, c'est-à-dire $\pi_T = (Id_X \times T) * \mu$. Alors, on a bien :

$$\int_X c(x, T(x)) \, d\mu(x) = \int_{X \times Y} c(x, y) \, d\pi_T(x, y).$$

L'existence d'une application de transport fournit un plan de transport admissible pour le problème (K). En revanche un plan de transport optimal pour le problème (K) n'est pas forcément issu d'une application de transport puisqu'une telle application n'existe pas toujours.

Interprétation du problème de Kantorovich : Reprenons l'exemple avec la masse de Dirac donné dans la partie sur le problème de Monge. Bien qu'aucune application de transport n'existe, un moyen de transporter δ_0 sur $\frac{\delta_0+\delta_1}{2}$ est de laisser la moitié de la masse en 0 et d'envoyer l'autre moitié sur 1. L'idée est donc de s'autoriser à répartir la masse présente en x dans le trou selon une probabilité λ_x, et ce pour tout x. Rappelons le théorème de Jirina :

Théorème 1.2. *Soient X et Y des espaces polonais et $\pi \in \Pi(\mu, \nu)$, alors il existe une application $\lambda : X \times Y \to [0, 1]$ telle que :*

- $\forall x \in X, \quad \lambda(x, .) \in \mathcal{P}(Y)$
- $\forall B \in \mathcal{B}(Y), \quad \lambda(. , B)$ *est $\mathcal{B}(X)$-mesurable*

- $\forall A \in \mathcal{B}(X), \quad \forall B \in \mathcal{B}(Y), \quad \pi(A \times B) = \displaystyle\int_A \lambda(x, B)\, d\mu(x).$

λ est appelé noyau de transition entre μ et ν.

L'interprétation de $\lambda(x, .) := \lambda_x$ est la répartition dans le trou de la masse de sable présente dans le tas de sable en x. En effet, on peut écrire : $d\pi(x, y) = d\lambda_x(y) \otimes d\mu(x)$.

Si toute la masse présente en x est envoyée sur $T(x)$, alors le noyau de transition est : $\lambda(x, .) = \delta_{T(x)}$ et on retrouve l'expression du coût dans le problème de Monge :

$$\int_{X \times Y} c(x, y)\, d\pi(x, y) = \int_X \left(\int_Y c(x, y)\, d\lambda_x(y) \right) d\mu(x) = \int_X c(x, T(x))\, d\mu(x).$$

1.2 Résolution du problème de Kantorovich

1.2.1 Topologie de la convergence étroite de probabilités

Rappelons quelques faits sur la notion de convergence étroite de probabilités. On se place sur un espace métrique (X, d).

Définition 1.3. *Soient $(\mu_n)_n$ une suite de $\mathcal{P}(X)$ et $\mu \in \mathcal{P}(X)$. On dit que la suite $(\mu_n)_n$ converge étroitement vers μ si :*

$$\forall \phi \in C_b(X, \mathbb{R}), \quad \int_X \phi\, d\mu_n \underset{n \to +\infty}{\to} \int_X \phi\, d\mu,$$

ce qu'on note : $\mu_n \rightharpoonup \mu$.

Remarque 1.4. *Il y a bien unicité de la limite. Supposons qu'une suite admette deux limites μ et ν pour la convergence étroite. Alors on a :*

$$\forall \phi \in C_b(X, \mathbb{R}), \quad \int_X \phi\, d\mu = \int_X \phi\, d\nu.$$

Par caractérisation de la loi par intégration contre une fonction continue bornée, on a bien $\mu = \nu$.

Rappelons une caractérisation utile de la convergence étroite dans le cas où $X = \mathbb{R}$ (avec la métrique induite par la valeur absolue).

Théorème 1.5. *Soient $(\mu_n)_n$ une suite de $\mathcal{P}(\mathbb{R})$ et $\mu \in \mathcal{P}(\mathbb{R})$. On note F_n la fonction de répartition de μ_n et F celle de μ. Il y a équivalence entre :*

1. *$\mu_n \rightharpoonup \mu$.*
2. *Pour tout point de continuité x de F, on a :*

$$F_n(x) \to F(x).$$

Enfin, nous admettons qu'il existe une distance qui métrise la convergence étroite des probabilités.

Théorème 1.6. *Soit (X, d) un espace polonais, il existe une distance sur $\mathcal{P}(X)$ qui métrise la convergence étroite.*

Pour avoir des détails à ce sujet, on pourra consulter [3].

1.2.2 Le théorème de Prokhorov

L'élément clé pour résoudre le problème de Kantorovich est le théorème de Prokhorov. Il s'agit d'un résultat de compacité dans l'espace des probabilités. Définissons d'abord le caractère tendu d'une famille de probabilités.

Définition 1.7. *Soit X un espace polonais. On dit qu'une partie $M \subset \mathcal{P}(X)$ est tendue si :*

$$\forall \epsilon > 0, \exists K \subset X \text{ compact}, \forall \mu \in M, \quad \mu(X \backslash K) \leq \epsilon.$$

On peut maintenant énoncer le théorème.

Théorème 1.8 (Théorème de Prokhorov)**.**
Soient (X, d) un espace polonais et $M \subset \mathcal{P}(X)$. M est relativement compact si et seulement si il est tendu.

Nous allons seulement démontrer ce théorème dans le cas $X = \mathbb{R}$, pour une preuve dans le cas général, on pourra consulter [3]. Nous allons avoir besoin plusieurs lemmes.

Lemme 1.9 (Lemme d'Ulam)**.**
Soit X un espace polonais et $\mu \in \mathcal{P}(X)$, alors $\{\mu\}$ est tendue.

Ce résultat est évident sur $\mathbb{R}^p$ puisque les fermés bornés sont compacts. On ne fera pas la preuve dans le cas général.

Lemme 1.10. *Soit $(\mu_n)_n$ une suite $\mathcal{P}(\mathbb{R}^p)$ telle que $\mu_n \rightharpoonup \mu$, où $\mu \in \mathcal{P}(\mathbb{R}^p)$. Alors la suite $(\mu_n)_n$ est tendue.*

Démonstration. Soit $\epsilon > 0$. Grâce au lemme 1.9 d'Ulam, on dispose de $A > 0$ tel que :

$$\mu(B(0, A)^c) \leq \frac{\epsilon}{2}.$$

On considère une fonction f, à valeurs dans $[0, 1]$, continue et affine par morceaux valant 1 sur $[-2A, 2A]^c$ et valant 0 sur $[-A, A]$ et on pose $h = f \circ |.| \in C_b(\mathbb{R}^p)$, où $|.|$ désigne la norme euclidienne canonique sur $\mathbb{R}^p$. On a :

$$\mathbf{1}_{B(0,2A)^c} \leq h \leq \mathbf{1}_{B(0,A)^c}.$$

D'après la seconde inégalité et par convergence étroite, on dispose de $n_0 \in \mathbb{N}$, tel que :

$$\forall n \geq n_0, \quad \int_{\mathbb{R}^p} h \, d\mu_n \leq \epsilon.$$

La première inégalité donne :

$$\forall n \geq n_o, \quad \mu_n(B(0, 2A)^c) \leq \int_{\mathbb{R}^p} h \, d\mu_n \leq \epsilon.$$

Puisque d'après le lemme 1.9 d'Ulam, un nombre fini de probabilités sur $\mathbb{R}^p$ est tendu, on peut agrandir le compact et trouver un compact K tel que :

$$\forall n \in \mathbb{N}, \quad \mu_n(K) \leq \epsilon.$$

$\square$

Théorème 1.11 (Théorème de Helly).

Soit $(F_n)_n$ une suite de fonctions de répartition. Alors il existe une extraction $(\alpha_n)_n$ et une fonction $F : \mathbb{R} \to [0,1]$ croissante et continue à droite telle qu'en tout point de continuité x de F, on a :

$$F_{\alpha_n}(x) \underset{n\to\infty}{\to} F(x).$$

Démonstration.

Soit $q \in \mathbb{Q}$, la suite $(F_n(q))_n$ est une suite de $[0,1]$, on peut donc extraire une sous-suite qui converge vers un réel qu'on note $\tilde{F}(q) \in [0,1]$. Par extraction diagonale, il existe une extraction $(\alpha_n)_n$ telle que :

$$\forall q \in \mathbb{Q},\, F_{\alpha_n}(q) \underset{n\to\infty}{\to} \tilde{F}(q).$$

On peut alors définir, si $x \in \mathbb{R}$:

$$F(x) = \inf_{q\in\mathbb{Q},\, x<q} \{\tilde{F}(q)\}.$$

F est croissante par construction. Montrons qu'elle est continue à droite. Soit $x \in \mathbb{R}$ et $\epsilon > 0$. Il existe $q > x$ tel que $\tilde{F}(q) \leq F(x) + \epsilon$. Ainsi, si $y \in [x, q]$, on a :

$$F(x) \leq F(y) \leq \tilde{F}(q) \leq F(x) + \epsilon.$$

Soit maintenant x un point de continuité de F et $\epsilon > 0$. Par continuité, il existe $y < x$ tel que :

$$F(x) - \epsilon \leq F(y).$$

Soient $r, s \in \mathbb{Q}$ tels que : $y < r < x < s$ et $\tilde{F}(s) \leq F(x) + \epsilon$. On a alors :

$$F(x) - \epsilon \leq F(y) \leq \tilde{F}(r) \leq \tilde{F}(s) \leq F(x) + \epsilon.$$

Or par définition, on a :

$$\tilde{F}(r) = \lim_n F_{\alpha_n}(r) \leq \underline{\lim}\, F_{\alpha_n}(x) \leq \overline{\lim}\, F_{\alpha_n}(x) \leq \overline{\lim}\, F_{\alpha_n}(s) = F(s).$$

En laissant tendre ϵ vers 0, on obtient $F_{\alpha_n}(x) \underset{n\to\infty}{\to} F(x)$. □

Démonstration. (Théorème de Prokhorov dans le cas $X = \mathbb{R}$).

($\Leftarrow$) Soit M un ensemble tendu de probabilités sur $\mathbb{R}$. Comme la convergence étroite est métrisable, il suffit de montrer la relative compacité séquentielle. Soit $(\mu_n)_n \in M^{\mathbb{N}}$. On note $(F_n)_n$ la suite de leur fonction de répartition. D'après le théorème 1.11 de Helly, il existe une extraction $(\alpha_n)_n$ et une fonction $F : \mathbb{R} \to [0,1]$ croissante et continue à droite telle qu'en tout point de continuité x de F, on a :

$$F_{\alpha_n}(x) \underset{n\to\infty}{\to} F(x).$$

On admet (grâce au théorème d'extension de mesures de Carathéodory) qu'on peut construire une mesure μ à valeurs dans $[0,1]$ telle que :

$$\forall\, a \leq b \quad \mu(]a,b]) = F(b) - F(a).$$

Puisque la suite $(\mu_n)_n$ est tendue et que les points de discontinuité de F sont dénombrables (donc l'ensemble des points de continuité de F est non borné), on a :

$$\forall \epsilon > 0, \exists u < v \text{ des points de continuité de } F, \forall n \in \mathbb{N} \quad \mu_n(]u, v]) \geq 1 - \epsilon.$$

On a donc, puisqu'on a choisi des points de continuité de F :

$$\forall \epsilon > 0, \quad \mu(\mathbb{R}) \geq \mu(]u, v]) = F(v) - F(u) = \lim_{n\to\infty} F_{\alpha_n}(v) - F_{\alpha_n}(u) \geq 1 - \epsilon.$$

En laissant tendre ϵ vers 0, on obtient $\mu \in \mathcal{P}(\mathbb{R})$ et $\mu_{\alpha_n} \rightharpoonup \mu$ puisque la suite des fonctions de répartition associées converge simplement vers F en ses points de continuité.

($\Rightarrow$) On raisonne par contraposition : on suppose que M n'est pas tendu. Il existe donc $\epsilon > 0$ et une suite $(\mu_n)_n \in M^{\mathbb{N}}$ tels que :

$$\forall n \in \mathbb{N}, \quad \mu_n([-n, n]) \leq 1 - \epsilon.$$

On considère une extraction quelconque $(\mu_{\alpha_n})_n$. Cette suite extraite n'est pas tendue. En effet, on a :

$$\forall n \in \mathbb{N}, \quad \mu_{\alpha_n}([-\alpha_n, \alpha_n]) \leq 1 - \epsilon$$

Puisque $\alpha_n \to +\infty$, il n'existe pas de compact K tel que :

$$\forall n \in \mathbb{N}, \quad \mu_{\alpha_n}(K) \geq 1 - \frac{\epsilon}{2}.$$

On suppose que cette suite extraite $(\mu_{\alpha_n})_n$ converge étroitement vers μ, le lemme 1.10 assure que la suite est tendue, ce qui est une contradiction.

$\square$

1.2.3 Existence de plans de transport optimaux

On peut maintenant passer à la résolution du problème de Kantorovich. L'existence d'un plan de transport optimal se fera sous une hypothèse relativement faible de la fonction coût.

Définition 1.12. *Soit (X, d) un espace métrique. Une application $f : X \to \mathbb{R}$ est dite semi-continue inférieurement (qu'on abrège par sci) si pour tout $x \in X$ et pour toute suite $(x_n)_n$ de X qui tend vers x on a :*

$$\underline{\lim}\, f(x_n) \geq f(x).$$

On énonce précisément le théorème.

Théorème 1.13. *Soient X et Y deux espaces polonais, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ et $c : X \times Y \to \mathbb{R}^+$ une fonction sci. Alors le problème (K) de Kantorovich admet une solution, c'est-à-dire un plan de transport optimal.*

Une première étape pour démontrer ce théorème est la compacité de l'ensemble $\Pi(\mu, \nu)$.

Théorème 1.14. *Soient X, Y des espaces polonais, $\mu \in \mathcal{P}(X)$ et $\nu \in \mathcal{P}(Y)$. Alors $\Pi(\mu, \nu)$ est compact.*

Démonstration. Montrons déjà que l'ensemble $\Pi(\mu,\nu)$ est fermé. Soit $(\pi_n)_n$ une suite de $\Pi(\mu,\nu)$ qui converge étroitement vers $\pi \in \mathcal{P}(X \times Y)$. Si $\phi \in C_b(X)$ et $\psi \in C_b(Y)$, on a :

$$\begin{aligned}
\int_{X\times Y} (\phi(x)+\psi(y))\, d\pi(x,y) &= \lim_{n\to+\infty} \int_{X\times Y} (\phi(x)+\psi(y))\, d\pi_n(x,y) \\
&= \lim_{n\to+\infty} \int_X \phi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \qquad \text{car } \pi_n \in \Pi(\mu,\nu) \\
&= \int_X \phi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y).
\end{aligned}$$

Les lois marginales de π sont donc μ et ν. Ainsi $\pi \in \Pi(\mu,\nu)$ et l'ensemble $\Pi(\mu,\nu)$ est donc fermé (pour la topologie de la convergence étroite).

Pour montrer que $\Pi(\mu,\nu)$ est compact, il suffit de montrer qu'il est tendu. En effet le théorème 1.8 de Prokhorov assurera qu'il est relativement compact et donc compact.

Soit $\epsilon > 0$. D'après le lemme 1.9 d'Ulam, on a l'existence de K compact de X et de L compact de Y tels que : $\mu(X\backslash K) \leq \epsilon$ et $\nu(Y\backslash L) \leq \epsilon$. Le compact $K \times L$ de $X \times Y$ vérifie :

$$\forall \pi \in \Pi(\mu,\nu), \quad \pi(X \times Y \setminus K \times L) \leq \mu(X\backslash K) + \nu(Y\backslash L) \leq 2\epsilon.$$

D'où la tension de $\Pi(\mu,\nu)$.

□

Il nous manque seulement le lemme suivant pour résoudre le problème de Kantorovich :

Lemme 1.15. *Soit (X,d) un espace polonais et f une fonction positive et sci sur X. Soit $(\mu_n)_n \in \mathcal{P}(X)^{\mathbb{N}}$ et $\mu \in \mathcal{P}(X)$ tel que $\mu_n \rightharpoonup \mu$, alors on a :*

$$\int_X f\, d\mu \leq \underline{\lim} \int_X f\, d\mu_n.$$

Démonstration. L'idée de la preuve est d'approcher la fonction f par une suite croissante de fonctions continue bornées. On pose ainsi, si $k \in \mathbb{N}$:

$$f_k : \left\{ \begin{array}{ccc} X & \to & \mathbb{R}^+ \\ x & \mapsto & \min\left\{ \inf\limits_{y\in X} \{f(y) + k\, d(x,y)\}\ ,\ k \right\} \end{array} \right. .$$

L'application f_k est continue car k-lipschitzienne comme inf de fonctions positives qui le sont. Montrons que la suite $(f_k)_k$ converge simplement vers f. Soit $x \in X$. Par définition, il existe une suite $(y_n)_n$ telle que :

$$\forall n > 0, \quad f_n(x) + \frac{1}{n} \geq f(y_n) + n\, d(y_n, x).$$

En prenant $y = x$ dans la définition de f_k, on obtient que pour tout $k \in \mathbb{N}$, $f_k \leq f$. De plus, par positivité de f, on a :

$$\forall n \in \mathbb{N}, \quad f(x) + \frac{1}{n} \geq f_n(x) + \frac{1}{n} \geq n\, d(y_n, x).$$

En laissant tendre n vers $+\infty$ on déduit : $y_n \to x$. Or on a également l'inégalité :

$$\forall n \in \mathbb{N}, \quad f(x) + \frac{1}{n} \geq f_n(x) + \frac{1}{n} \geq f(y_n).$$

En prenant la limite inférieure, on a :

$$f(x) \geq \underline{\lim}\, f_n(x) \geq \underline{\lim}\, f(y_n) \geq f(x),$$

la dernière inégalité venant de la semi-continuité inférieure de f. La suite $(f_n(x))_n$ étant croissante et majorée par f, on obtient la convergence de $(f_n)_n$ vers f.

Le théorème de Beppo-Levi assure alors que :

$$\int_X f_k \, d\mu \underset{k \to +\infty}{\to} \int_X f \, d\mu.$$

Or par convergence étroite, on a :

$$\forall k \in \mathbb{N}, \quad \int_X f_k \, d\mu_n \underset{n \to +\infty}{\to} \int_X f_k \, d\mu.$$

Comme $\forall k \in \mathbb{N} \quad f_k \leq f$, on déduit (en passant à la limite inférieure) :

$$\int_X f_k \, d\mu = \underline{\lim}_n \int_X f_k \, d\mu_n \leq \underline{\lim}_n \int_X f \, d\mu_n.$$

En passant à la limite quand $k \to +\infty$, on obtient l'inégalité voulue. $\square$

Démontrons maintenant le théorème 1.13 de Kantorovich.

Démonstration. On prend une suite minimisante $(\pi_k)_k$ de $\Pi(\mu, \nu)$ dans le problème de Kantorovich :

$$\int_{X \times Y} c \, d\pi_k \underset{k \to +\infty}{\to} \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c \, d\pi.$$

Par compacité de $\Pi(\mu, \nu)$, il existe une extraction, toujours notée $(\pi_k)_k$ qui converge étroitement vers $\pi \in \Pi(\mu, \nu)$.

D'après le lemme 1.15, on a :

$$\int_{X \times Y} c \, d\pi \leq \underline{\lim} \int_{X \times Y} c \, d\pi_k = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c \, d\pi.$$

$\square$

On a donc montré qu'il existe une solution au problème de Kantorovich, qu'on appelle plan de transport optimal.

1.2.4 Dualité de Kantorovich-Rubinstein

Dans cette partie on présente rapidement la notion de problème dual du problème de Kantorovich. On se donne toujours X et Y deux espaces polonais, ainsi que $\mu \in \mathcal{P}(X)$ et $\nu \in \mathcal{P}(Y)$. On définit l'ensemble

$$\phi_c := \left\{ (\phi, \psi) \in L^1(Y,\nu) \times L^1(X,\mu), \quad \phi(y) + \psi(x) \leq c(x,y), \quad \mu\text{-pp } x \ , \ \nu\text{-pp } y \ \right\}.$$

De plus, si $(\phi, \psi) \in \phi_c$, on définit :

$$J(\phi, \psi) := \int_Y \phi \, d\nu + \int_X \psi \, d\mu.$$

On admet le théorème général de dualité de Kantorovich, dont on pourra trouver une preuve dans [2].

Théorème 1.16 (Théorème de dualité de Kantorovich)**.**
On se place dans le même cadre que le théorème 1.13 de Kantorovich. Alors on a :

$$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \sup_{(\phi,\psi) \in \phi_c} J(\phi, \psi) = \sup_{(\phi,\psi) \in \phi_c \cap C_b} J(\phi, \psi).$$

Énonçons maintenant un cas particulier de ce théorème dans le cas où la fonction coût est donnée par une distance.

Théorème 1.17 (Dualité de Kantorovich-Rubinstein)**.**
Soit X un espace polonais et d une distance sci sur X. On se donne μ et ν dans $\mathcal{P}(X)$. Alors :

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} d(x,y) \, d\pi(x,y) = \sup_{\|\phi\|_{lip} \leq 1, \phi \text{ bornée}} \int_X \phi \, d\mu - \int_Y \phi \, d\nu.$$

Démonstration. Soit ϕ une fonction bornée et lipschitzienne avec :

$$\|\phi\|_{lip} := \sup_{x \neq y} \left\{ \frac{|\phi(x) - \phi(y)|}{d(x,y)} \right\} \leq 1.$$

Soit $\pi \in \Pi(\mu, \nu)$. On a :

$$\int_X \phi \, d\mu - \int_Y \phi \, d\nu = \int_{X \times Y} \phi(x) - \phi(y) \, d\pi(x,y) \leq \int_{X \times Y} d(x,y) \, d\pi(x,y).$$

D'où : $\displaystyle\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} d(x,y) \, d\pi(x,y) \geq \sup_{\|\phi\|_{lip} \leq 1, \phi \text{ bornée}} \int_X \phi \, d\mu - \int_Y \phi \, d\nu.$

Pour avoir l'autre inégalité, on utilise le théorème 1.16 de dualité de Kantorovich :

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} d(x,y) \, d\pi(x,y) = \sup_{\substack{\phi, \psi \in C_b \\ \phi(y) + \psi(x) \leq d(x,y)}} \int_X \phi \, d\nu + \int_X \psi \, d\mu.$$

Or l'inégalité de contrainte assure que pour tout $x, y \in X$:

$$\phi(y) \leq d(x,y) - \psi(x).$$

Puis en prenant l'inf en x :

$$\phi(y) \leq \inf_{x\in X} \{d(x,y) - \psi(x)\} := \psi^d(y).$$

L'application ψ^d ainsi définie est 1-lipschitzienne comme inf de fonctions qui le sont. Elle est également bornée car pour tout $y \in X$, on a : $\psi^d(y) \leq -\psi(y)$. D'où :

$$J(\phi,\psi) \leq J(\psi^d,\psi).$$

De plus, on obtient successivement les inégalités :

$$\begin{aligned} \psi^d(y) &\leq d(x,y) - \psi(x) \quad \forall x,y \in X \\ \psi(x) &\leq d(x,y) - \psi^d(y) \quad \forall x,y \in X. \end{aligned}$$

Puis en prenant l'inf en $y \in X$:

$$\psi(x) \leq \psi^{dd}(x) \quad \forall x \in X.$$

On a donc montré que : $J(\psi^d,\psi) \leq J(\psi^d,\psi^{dd})$.

Finalement, on a :

$$\sup_{\substack{\phi,\psi\in C_b \\ \phi(y)+\psi(x)\leq d(x,y)}} J(\phi,\psi) \leq \sup_{\psi\in C_b} J(\psi^d,\psi^{dd}) \leq \sup_{\substack{\phi,\psi\in C_b \\ \phi(y)+\psi(x)\leq d(x,y)}} J(\phi,\psi).$$

D'où l'égalité entre ces deux quantités.

Pour achever la preuve, nous allons montrer que $\psi^{dd} = -\psi^d$. Par définition de ψ^{dd}, en prenant $y = x$, on a : $\psi^{dd} \leq -\psi^d$. De plus, on a successivement :

$$\begin{aligned} \psi^d(y) - \psi^d(x) &\geq -d(x,y) \quad \forall x,y \in X \\ -\psi^d(y) &\leq d(x,y) - \psi^d(x) \quad \forall x,y \in X. \end{aligned}$$

D'où :

$$-\psi^d(y) \leq \inf_{x\in X} \left\{ d(x,y) - \psi^d(x) \right\} = \psi^{dd}(y), \quad \forall y \in X.$$

L'égalité est donc prouvée. Pour conclure, on écrit :

$$\sup_{\psi\in C_b} J(\psi^d,\psi^{dd}) = \sup_{\psi\in C_b} J(\psi^d,-\psi^d) \leq \sup_{\|\phi\|_{lip}\leq 1, \phi \text{ bornée}} J(\phi,-\phi).$$

□

1.3 Résolution du problème de Monge dans un cas particulier : le théorème de Brenier

Nous allons résoudre le problème de Monge dans un cas particulier, c'est-à-dire que le plan de transport optimal du problème de Kantorovich est en fait issu d'une application de transport.

Théorème 1.18 (Théorème de Brenier).

On se place sur $\mathbb{R}^n$ muni de sa norme euclidienne canonique notée $|.|$. Soient $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$, dont les moments d'ordre 2 sont finis, i.e :

$$\int_{\mathbb{R}^n} |x|^2 \, d\mu(x) + \int_{\mathbb{R}^n} |x|^2 \, d\nu(x) < +\infty.$$

On suppose de plus que μ est absolument continue par rapport à la mesure de Lebesgue. On se place dans le cas où la fonction coût de transport est donnée par :

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

Alors il existe un unique plan de transport optimal π, qui est en fait déterministe, c'est-à-dire donné par une application de transport T transportant μ sur ν avec :

$$d\pi(x, y) = \delta_{T(x)}(y) \otimes d\mu(x).$$

De plus, l'application T s'écrit en fait comme le gradient (défini de manière unique μ-pp) d'une fonction convexe

$$\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}.$$

La preuve de ce théorème se trouve dans [1].

2 Distances de Wasserstein

Le problème de transport de Kantorovich permet de quantifier l'écart entre deux mesures de probabilités selon une fonction coût donnée. Typiquement la distance sur X peut être prise comme fonction coût dans le cas où $X = Y$. On se demande dans quelle mesure cela permet de définir une distance sur l'espace des probabilités. Nous allons nous restreindre aux sous-espaces des probabilités ayant un moment d'ordre p fini pour y définir une distance. Dans la suite, (X, d) désigne un espace polonais et $p \in [1, +\infty)$.

2.1 Définition des distances de Wasserstein W_p, avec $p \geq 1$

Définition 2.1. *Soit $x_0 \in X$. On définit :*

$$\mathcal{P}_p(X) = \left\{ \mu \in \mathcal{P}(X), \quad \int_X d(x, x_0)^p \, d\mu(x) < +\infty \right\}.$$

Si $\mu \in \mathcal{P}_p(X)$, on dira que μ admet un moment d'ordre p fini.

Remarque 2.2.

- *Il est facile de voir, grâce à l'inégalité triangulaire, que si la condition dans la définition est réalisée pour un x_0, alors elle l'est pour tout x_0.*
- *Si d est une distance bornée, alors pour tout $p \geq 1$, on a : $\mathcal{P}_p(X) = \mathcal{P}(X)$.*
- *La définition de ces espaces est naturelle dans le cadre du transport d'une probabilité $\mu \in \mathcal{P}(X)$ sur la mesure de Dirac δ_{x_0}. En effet, on a :*

$$\inf_{\pi \in \Pi(\mu, \delta_{x_0})} \int_{X \times X} d(x, y)^p \, d\pi(x, y) = \int_X d(x, x_0)^p \, d\mu(x)$$

puisque $\Pi(\mu, \delta_{x_0}) = \{\mu \otimes \delta_{x_0}\}$.

Théorème 2.3. *L'application W_p définie par :*

$$W_p : \begin{cases} \mathcal{P}_p(X) \times \mathcal{P}_p(X) & \to \quad \mathbb{R}^+ \\ (\mu, \nu) & \mapsto \quad \left(\inf\limits_{\pi \in \Pi(\mu,\nu)} \displaystyle\int\limits_{X \times X} d(x,y)^p \, d\pi(x,y) \right)^{1/p} \end{cases}$$

est une distance sur $\mathcal{P}_p(X)$, appelée distance de Wasserstein.

Remarque 2.4. *Une réécriture plus concise de la distance de Wasserstein est la suivante :*

$$\forall \mu, \nu \in \mathcal{P}_p(X), \quad W_p(\mu,\nu) = \left(\inf_{U \sim \mu,\, V \sim \nu} \mathbb{E}\, d(U,V)^p \right)^{1/p},$$

où on note $U \sim \mu$ une variable aléatoire de loi μ.

Exemple : Si on se place sur $\mathbb{R}^n$ muni de la distance euclidienne classique, et si $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ avec X une variable aléatoire de loi μ, alors :

$$W_2^2(\mu, \delta_{\mathbb{E}(X)}) = \mathrm{Var}(X) := \int_{\mathbb{R}^n} \left| x - \int_{\mathbb{R}^n} x \, d\mu(x) \right|^2 d\mu(x).$$

Avant de démontrer le théorème 2.3, énonçons un lemme que nous utiliserons dans la preuve de l'inégalité triangulaire.

Lemme 2.5 (Lemme de recollement)**.**
Soient X_1, X_2, X_3 des espaces polonais, μ_1, μ_2 et μ_3 des probabilités respectivement sur X_1, X_2 et X_3 et soient enfin $\pi_{1,2} \in \Pi(\mu_1, \mu_2)$ et $\pi_{2,3} \in \Pi(\mu_2, \mu_3)$. Alors il existe $\pi \in \mathcal{P}(X_1 \times X_2 \times X_3)$ de marges $\pi_{1,2}$ sur $X_1 \times X_2$ et $\pi_{2,3}$ sur $X_2 \times X_3$.

La preuve de ce théorème repose sur le théorème 1.2 de désintégration de Jirina. Nous allons maintenant démontrer le théorème 2.3.

Démonstration.

- La quantité définie est à valeurs dans $\mathbb{R}^+$.

  La positivité est immédiate. Soient $\mu, \nu \in \mathcal{P}_p(X)$ et $\pi \in \Pi(\mu,\nu)$. Alors :

$$\begin{aligned} \int\limits_{X \times X} d(x,y)^p \, d\pi(x,y) &\leq \int\limits_{X \times X} (d(x,x_0) + d(x_0,y))^p \, d\pi(x,y) \quad \text{par inégalité triangulaire} \\ &\leq \int\limits_{X \times X} (2\, max(d(x,x_0), d(y,x_0)))^p \, d\pi(x,y) \\ &\leq 2^p \left(\int_X d(x,x_0)^p \, d\mu(x) + \int_X d(y,x_0)^p \, d\nu(y) \right) \\ &< +\infty \quad \text{car } \mu, \nu \in \mathcal{P}_p(X). \end{aligned}$$

- Symétrie.

  Soient $\mu, \nu \in \mathcal{P}_p(X)$. Il suffit de remarquer que :

$$\begin{aligned} W_p^p(\mu,\nu) &= \inf_{\pi\in\Pi(\mu,\nu)} \int\limits_{X\times X} d(x,y)^p \, d\pi(x,y) \\ &= \inf_{\pi\in\Pi(\mu,\nu)} \int\limits_{X\times X} d(y,x)^p \, d\pi(x,y) \\ &= W_p^p(\nu,\mu). \end{aligned}$$

- Séparation

  On remarque déjà que si $\mu \in \mathcal{P}_p(X)$, alors $W_p(\mu,\mu) = 0$ en prenant $\pi = (Id_X \times Id_X) * \mu$.

  Soient $\mu, \nu \in \mathcal{P}_p(X)$ telles que $W_p(\mu,\nu) = 0$. Ceci implique qu'il existe un plan de transport optimal π qui est concentré sur la diagonale $D := \{(x,x),\, x \in X\}$. Alors si $A \in \mathcal{B}(X)$:

$$\begin{aligned} \mu(A) = \pi(A\times X) = \pi(A\times X \cap D) &= \pi((x,y),\, x\in A,\, y=x) = \pi((x,y),\, y\in A,\, x=y) \\ &= \pi(X\times A) = \nu(A). \end{aligned}$$

  D'où $\mu = \nu$.

- Inégalité triangulaire.

  Rédigeons la preuve avec la notation probabiliste introduite dans la remarque 2.4. On se donne $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_p(X)$ et $X_1 \sim \mu_1, X_2 \sim \mu_2, X_3 \sim \mu_3$ des variables aléatoires. D'après le théorème 1.13, il existe un plan optimal $\pi_{1,2}$ entre μ_1 et μ_2, et un plan optimal $\pi_{2,3}$ entre μ_2 et μ_3. Le lemme 2.5 de recollement assure qu'il existe $\pi \in \mathcal{P}(X^3)$ qui recolle $\pi_{1,2}$ et $\pi_{2,3}$. Si $(Y_1,Y_2,Y_3) \sim \pi$, alors $(Y_1,Y_2) \sim \pi_{1,2}$ et $(Y_2,Y_3) \sim \pi_{2,3}$. On a alors :

$$\begin{aligned} W_p(\mu_1,\mu_3) &\leq \left(\mathbb{E}\, d(Y_1,Y_3)^p\right)^{1/p} \quad \text{(par définition)} \\ &\leq \left(\mathbb{E}\,(d(Y_1,Y_2)+d(Y_2,Y_3))^p\right)^{1/p} \quad \text{(inégalité triangulaire sur } X\text{)} \\ &\leq \left(\mathbb{E}\, d(Y_1,Y_2)^p\right)^{1/p} + \left(\mathbb{E}\, d(Y_2,Y_3)^p\right)^{1/p} \quad \text{(inégalité de Minkowski)} \\ &= W_p(\mu_1,\mu_2) + W_p(\mu_2,\mu_3) \quad \text{(d'après le lemme de recollement).} \end{aligned}$$

□

L'inégalité de Hölder entraîne directement la proposition qui suit.

Proposition 2.6. *Soient $1 \leq p \leq q < +\infty$, alors : $W_p \leq W_q$.*

Inversement, si la distance d est bornée, alors : $W_q^q \leq (diam(X))^{q-p}\, W_p^p$.

2.2 Convergence en distance W_p et lien avec la convergence étroite

Nous allons maintenant nous intéresser à la convergence d'une suite de probabilités pour une distance de Wasserstein et au lien avec la convergence étroite. Énonçons d'abord un lemme et une proposition qui seront utiles dans la suite. On démontrera seulement la proposition.

Lemme 2.7. *Soit $(\mu_n)_n$ une suite de Cauchy de $(\mathcal{P}_p(X), W_p)$, alors $(\mu_n)_n$ est tendue.*

La preuve de ce lemme est basée sur l'utilisation du théorème 1.17 de Kantorovich-Rubinstein.

Proposition 2.8. *Soit $(\mu_n)_n \in \mathcal{P}_p(X)^{\mathbb{N}}$ qui converge étroitement vers $\mu \in \mathcal{P}(X)$. Alors :*

1. $\displaystyle\int_X d(x, x_0)^p \, d\mu(x) \leq \varliminf_n \int_X d(x, x_0)^p \, d\mu_n(x), \quad \forall x_0 \in X$
2. $\displaystyle\forall \nu \in \mathcal{P}_p(X), \qquad \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \int\limits_{X \times X} d(x, y)^p \, d\pi(x, y) \right\} \leq \varliminf_n W_p^p(\mu_n, \nu).$

   Ainsi si $\mu \in \mathcal{P}_p(X)$, alors :

$$W_p(\mu, \nu) \leq \varliminf_n W_p(\mu_n, \nu).$$

Démonstration. 1. Il s'agit du lemme 1.15.

2. On considère une extractrice ϕ qui réalise la limite inférieure :

$$W_p(\mu_{\phi(n)}, \nu) \underset{n \to +\infty}{\to} \varliminf_n W_p(\mu_n, \nu).$$

Or le théorème 1.13 assure que pour tout $n \in \mathbb{N}$, il existe $\pi_{\phi(n)} \in \Pi(\mu_{\phi(n)}, \nu)$ telle que :

$$W_p^p(\mu_{\phi(n)}, \nu) = \int_X d(x, y)^p \, d\pi_{\phi(n)}(x, y).$$

D'après le théorème 1.8 de Prokhorov et le lemme 1.9 d'Ulam, la famille $\{\mu_{\phi(n)}\} \cup \{\nu\}$ est tendue. Ainsi, si K est un compact de X, et si $n \in \mathbb{N}$, on a :

$$\pi_{\phi(n)}(X \times X \backslash K \times K) \leq \mu_{\phi(n)}(X \backslash K) + \nu(X \backslash K)$$

ce qui assure la tension de la famille $(\pi_{\phi(n)})_n$. Le théorème 1.8 de Prokhorov assure qu'il existe une extraction ψ et $\pi \in \mathcal{P}(X \times X)$ telles que :

$$\pi_{\phi(\psi(n))} \rightharpoonup \pi.$$

Alors on a :

$$\begin{aligned}
\int\limits_{X \times X} d(x, y)^p \, d\pi(x, y) &\leq \varliminf_n \int\limits_{X \times X} d(x, y)^p \, d\pi_{\phi(\psi(n))}(x, y) \quad \text{par 1.} \\
&= \varliminf_n W_p^p(\mu_{\phi(\psi(n))}, \nu) \\
&= \varliminf_n W_p^p(\mu_n, \nu).
\end{aligned}$$

Par convergence étroite de la première marge, on a bien $\pi \in \Pi(\mu, \nu)$, ce qui achève la preuve.

□

On peut maintenant s'intéresser au lien entre la convergence d'une suite pour la distance de Wasserstein W_p et la convergence étroite.

Définition 2.9. *On définit :*

$$C_{b,p} := \left\{ f \in C(X), (1 + d(x_0, .))^{-p} f \in C_b(X) \right\} \supset C_b(X).$$

Théorème 2.10. *Soient $(\mu_n)_n \in \mathcal{P}_p(X)^{\mathbb{N}}$ et $\mu \in \mathcal{P}_p(X)$. Alors, il y a équivalence entre :*

1. $\mu_n \overset{W_p}{\to} \mu$.
2. $\forall f \in C_{b,p}(X), \quad \displaystyle\int_X f \, d\mu_n \underset{n \to +\infty}{\to} \int_X f \, d\mu.$
3. $\mu_n \rightharpoonup \mu \quad$ *et* $\quad \displaystyle\int_X d(x, x_0)^p \, d\mu_n(x) \underset{n \to +\infty}{\to} \int_X d(x, x_0)^p \, d\mu(x), \quad \forall x_0 \in X.$
4. $\mu_n \rightharpoonup \mu \quad$ *et* $\quad \displaystyle\overline{\lim_n} \int_X d(x, x_0)^p \, d\mu_n(x) \leq \int_X d(x, x_0)^p \, d\mu(x), \quad \forall x_0 \in X.$
5. $\mu_n \rightharpoonup \mu \quad$ *et* $\quad \displaystyle\lim_{R \to +\infty} \overline{\lim_n} \int_X d(x, x_0)^p \, d\mu_n(x) \leq 0.$

Nous allons démontrer seulement une implication dans le théorème 2.10. La preuve des autres implications se trouve dans [1].

Démonstration. 1. $\Rightarrow$ 4.

Soit $(\mu_n)_n$ une suite de $\mathcal{P}_p(X)$ qui converge vers $\mu \in \mathcal{P}_p(X)$ en distance W_p. D'après le lemme 2.7, la suite $(\mu_n)_n$ est tendue, donc admet une sous-suite $(\mu_{\phi(n)})_n$ qui converge étroitement vers $\tilde{\mu} \in \mathcal{P}(X)$ d'après le théorème 1.8 de Prokhorov. D'après la proposition 2.8, on a :

$$W_p(\tilde{\mu}, \mu) \leq \underline{\lim_n} \, W_p(\mu_{\phi(n)}, \mu) = 0.$$

Ainsi $\mu = \tilde{\mu}$. En fait le raisonnement qu'on a fait peut se faire sur une sous suite quelconque de $(\mu_n)_n$. On déduit donc que :

$$\mu_n \rightharpoonup \mu$$

puisque de toute sous-suite on peut extraire une sous-suite convergeant étroitement vers μ.Il reste à montrer l'inégalité du point 4. On se donne $\epsilon > 0$. On remarque que :

$$\forall x, y \in \mathbb{R}^+, \quad (x + y)^p \leq (1 + \epsilon)^p x^p + \left(1 + \frac{1}{\epsilon}\right)^p y^p.$$

Pour démontrer cette inégalité, il suffit de discuter selon que $y \leq \epsilon x$ ou que $y > \epsilon x$. Dans le premier cas, le membre de gauche est majoré par le premier terme du membre de droite. Dans le second cas, on a la majoration avec le second terme du membre de droite. Il s'ensuit que pour tous $x, y \in X$:

$$\begin{aligned} d(x, x_0)^p &\leq (d(x, y) + d(y, x_0))^p \text{ par inégalité triangulaire} \\ &\leq (1 + \epsilon)^p d(y, x_0)^p + \left(1 + \frac{1}{\epsilon}\right)^p d(x, y)^p, \text{ d'après l'inégalité précédente.} \end{aligned}$$

On note π_n un plan de transport optimal envoyant μ sur μ_n, et on intègre l'inégalité obtenue précédemment contre la probabilité π_n :

$$\int_X d(x, x_0)^p \, d\mu(x) \leq (1 + \epsilon)^p \int_X d(x_0, y)^p \, d\mu_n(y) + \left(1 + \frac{1}{\epsilon}\right)^p \int_{X \times X} d(x, y)^p \, d\pi_n(x, y).$$

La deuxième intégrale à droite est égale à $W_p^p(\mu,\mu_n)$ qui tend vers 0 quand n tend vers $+\infty$. Ainsi :

$$\overline{\lim_n} \int_X d(x,x_0)^p \, d\mu(x) \leq (1+\epsilon)^p \int_X d(x_0,y)^p \, d\mu_n(y),$$

en passant à la limite supérieure. Il reste juste à laisser tendre ϵ vers 0 pour obtenir l'inégalité voulue.

□

Remarque 2.11. - *On remarque que si d est une distance bornée, alors $\mathcal{P}_p(X) = \mathcal{P}(X)$ et la convergence en distance de Wasserstein coïncide avec la convergence étroite puisque $C_b = C_{b,p}$.*

- *Dans le cas où d est une distance non bornée, si $\tilde{d}$ est une distance bornée définissant la même topologie que d (par exemple $\tilde{d} = \frac{d}{1+d}$), alors n'importe laquelle des distances de Wasserstein métrise la convergence étroite.*

2.3 Propriétés de complétude et de séparabilité

Le théorème essentiel est le suivant.

Théorème 2.12. *L'espace $(P_p(X), W_p)$ est un espace polonais.*

Démonstration.

- Montrons tout d'abord la complétude de l'espace (on a déjà vu au théorème 2.3 qu'il s'agit bien d'un espace métrique). On considère donc une suite de Cauchy $(\mu_n)_n$ dans $(P_p(X), W_p)$. D'après le lemme 2.7, la suite $(\mu_n)_n$ est tendue. Ainsi il existe une extractrice ϕ et $\mu \in \mathcal{P}(X)$ telles que :

$$\mu_{\phi(n)} \rightharpoonup \mu.$$

Montrons que la convergence a en fait lieu pour la distance W_p. D'après le théorème 1.13, pour tout $n, m \in \mathbb{N}$, il existe $\pi_{\phi(n),\phi(m)} \in \Pi(\mu_{\phi(n)}, \mu_{\phi(m)})$ un plan de transport optimal. m étant fixé, la suite $(\pi_{\phi(n),\phi(m)})_n$ est tendue d'après le même raisonnement que dans la preuve du point 2. de la proposition 2.8 et car la suite $(\mu_{\phi(n)})_n$ est elle-même tendue. Ainsi on peut extraire une sous-suite convergeant étroitement à m fixé :

$$\pi_{\phi(\psi(n)),\phi(m)} \underset{n\to+\infty}{\rightharpoonup} \pi_{\phi(m)} \in \mathcal{P}(X\times X).$$

Le lemme 1.15 assure alors qu'à m fixé :

$$\begin{aligned}\int_{X\times X} d(x,y)^p \, d\pi_{\phi(m)}(x,y) &\leq \underline{\lim_n} \int_{X\times X} d(x,y)^p \, d\pi_{\phi(\psi(n)),\phi(m)}(x,y) \\ &= \underline{\lim_n} W_p^p(\mu_{\phi(\psi(n))}, \mu_{\phi(m)}).\end{aligned}$$

Or puisque $\pi_{\phi(\psi(n)),\phi(m)} \in \Pi(\mu_{\phi(\psi(n))}, \mu_{\phi(m)})$, on en déduit que $\pi_{\phi(m)} \in \Pi(\mu, \mu_{\phi(m)})$ par convergence étroite de la suite $(\mu_{\phi(\psi(n))})_n$ vers μ.

Combinant les deux résultats précédents, on a :

$$\forall m \in \mathbb{N}, \quad W_p^p(\mu, \mu_{\phi(m)}) \leq \underline{\lim_n} W_p^p(\mu_{\phi(\psi(n))}, \mu_{\phi(m)}).$$

D'après le critère de Cauchy, le terme de droite est arbitrairement petit si m est suffisamment grand, ce qui assure que :

$$\mu_{\phi(m)} \xrightarrow{W_p} \mu.$$

On a donc trouvé une valeur d'adhérence à une suite de Cauchy, cette dernière converge donc vers μ en distance de Wasserstein W_p.

- Donnons simplement l'idée de la preuve de la séparabilité. On considère une suite $(x_n)_n$ dense dans (X, d). Il suffit de montrer que l'ensemble

$$\left\{ \mu = \sum_{n=1}^{N} b_n \delta_{x_n},\ N \geq 1,\ b_n \in \mathbb{Q}^+,\ \sum_{n \geq 1} b_n = 1 \right\}$$

est dense dans $(\mathcal{P}_p(X), W_p)$.

□

2.4 Convexité de W_p^p

L'inégalité suivante, dite de convexité, s'avérera être très utile dans la suite pour obtenir certaines majorations. Commençons par le cas le plus simple du cas d'une combinaison convexe de deux mesures.

Théorème 2.13. *Soient $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{P}_p(X)$ et $\alpha \in [0, 1]$. Alors on a l'inégalité dite de convexité suivante :*

$$W_p^p(\alpha\mu_1 + (1-\alpha)\mu_2, \alpha\nu_1 + (1-\alpha)\nu_2) \leq \alpha W_p^p(\mu_1, \nu_1) + (1-\alpha) W_p^p(\mu_2, \nu_2)$$

Démonstration. On prend un plan de transport optimal π_1 envoyant μ_1 sur ν_1 et un plan optimal π_2 envoyant μ_2 sur ν_2. On considère la probabilité :

$$\pi = \alpha\pi_1 + (1-\alpha)\pi_2 \in \mathcal{P}(X \times X).$$

On remarque que les marges de π sont $\alpha\mu_1 + (1-\alpha)\mu_2$ et $\alpha\nu_1 + (1-\alpha)\nu_2$. Ainsi on a :

$$\begin{aligned} W_p^p(\alpha\mu_1 + (1-\alpha)\mu_2, \alpha\nu_1 + (1-\alpha)\nu_2) &\leq \int\limits_{X \times X} d(x,y)^p \, d\pi(x,y) \\ &= \alpha \int\limits_{X \times X} d(x,y)^p \, d\pi_1(x,y) + (1-\alpha) \int\limits_{X \times X} d(x,y)^p \, d\pi_2(x,y) \\ &= \alpha W_p^p(\mu_1, \nu_1) + (1-\alpha) W_p^p(\mu_2, \nu_2). \end{aligned}$$

□

Remarque 2.14. *Si $\pi := \alpha\pi_1 + (1-\alpha)\pi_2$ est un plan de transport optimal, alors il y a égalité.*

Nous allons maintenant généraliser l'inégalité de convexité au cas où les probabilités ne sont plus combinaisons convexes de deux probabilités, mais définies par intégration d'une famille mesurable de probabilités (cf la définition 4.2 en annexe).

Avant cela, énonçons un théorème dont nous aurons besoin.

Théorème 2.15 (Sélection mesurable de plans de transports optimaux)**.**
Soient X et Y deux espaces polonais et $c : X \times Y \mapsto \mathbb{R}$ une application continue positive. Soient également T un espace mesurable et une application mesurable :

$$\left\{ \begin{array}{ccc} T & \to & \mathcal{P}(X) \times \mathcal{P}(Y) \\ t & \mapsto & (\mu(t), \nu(t)) \end{array} \right. .$$

Alors il existe une application mesurable :

$$\left\{ \begin{array}{ccc} T & \to & \mathcal{P}(X \times Y) \\ t & \mapsto & \pi(t) \end{array} \right.$$

telle que pour tout t, $\pi(t)$ est un plan de transport optimal envoyant $\mu(t)$ sur $\nu(t)$.

La preuve de ce théorème se trouve dans [2] à la page 91.

Théorème 2.16. *On suppose ici (X, d) localement compact. Soient (T, λ) un espace probabilisé et une application mesurable :*

$$\left\{ \begin{array}{ccc} T & \to & \mathcal{P}_p(X) \times \mathcal{P}_p(X) \\ t & \mapsto & (\mu(t), \nu(t)) \end{array} \right. .$$

Alors on a l'inégalité de convexité :

$$W_p^p \left(\int_T \mu(t) \, d\lambda(t), \int_T \nu(t) \, d\lambda(t) \right) \leq \int_T W_p^p(\mu(t), \nu(t)) \, d\lambda(t).$$

Démonstration. Pour tout $t \in T$, on dispose d'un plan de transport optimal $\pi(t)$ entre $\mu(t)$ et $\nu(t)$. D'après le théorème 2.15, on peut choisir $\pi(t)$ de telle sorte que $t \mapsto \pi(t)$ soit mesurable. On pose :

$$\pi := \int_T \pi(t) \, d\lambda(t),$$

qui est bien défini d'après le théorème 4.2 en annexe. On remarque que π a pour marges :

$$\int_T \mu(t) \, d\lambda(t) \quad \text{et} \quad \int_T \nu(t) \, d\lambda(t).$$

On a donc :

$$\begin{aligned} W_p^p \left(\int_T \mu(t) \, d\lambda(t), \int_T \nu(t) \, d\lambda(t) \right) &\leq \int_{X \times X} d(x,y)^p \, d\pi(x,y) \\ &= \int_T \left(\int_{X \times X} d(x,y)^p \, d\pi(t)(x,y) \right) d\lambda(t) \\ &= \int_T W_p^p(\mu(t), \nu(t)) \, d\lambda(t). \end{aligned}$$

□

3 Application à l'étude de l'équation de Boltzmann homogène en espace

3.1 L'équation de Boltzmann

On considère un gaz qu'on décrit par une densité de présence des particules de gaz $f(t,x,v)$ où t est le temps, $x \in \mathbb{R}^3$ la position d'une particule dans le gaz et $v \in \mathbb{R}^3$ la vitesse de la particule. On se restreint au cas où le gaz est homogène en espace, c'est-à-dire que la densité ne dépend pas de la position x. On considère qu'il n'y a pas de forces extérieures. Ainsi en l'absence de chocs, les particules évoluent en lignes droites. Seules les collisions modifient les trajectoires des particules : on parle d'équation cinétique.

On considère, dans le modèle de Boltzmann, des chocs élastiques (c'est-à-dire qui préservent la quantité de mouvement et l'énergie cinétique) impliquant uniquement deux particules (supposées de même masse). Notons v' et w' les vitesses des deux particules qui entrent en collision et v et w les vitesses après la collision. On suppose donc que :

$$v' + w' = v + w$$
$$|v'|^2 + |w'|^2 = |v|^2 + |w|^2.$$

En posant $v' = \frac{v+w}{2} + k_1$ et $w' = \frac{v+w}{2} - k_2$, les deux équations dont on dispose donnent :

$$k_1 = k_2$$
$$|k_1|^2 = \frac{|v-w|^2}{4}.$$

Ainsi il existe $\sigma \in S^2$, un paramètre de liberté, tel que :

$$v' = \frac{v+w}{2} + \frac{|v-w|}{2}\sigma$$
$$w' = \frac{v+w}{2} - \frac{|v-w|}{2}\sigma$$

En supposant l'indépendance des vitesses des deux particules qui entrent en collision (hypothèse dite de chaos moléculaire) et en faisant une hypothèse de type réversibilité au niveau microscopique, Boltzmann a montré que l'évolution de la quantité $f(t,v)$ est régie par l'équation :

$$\frac{\partial f}{\partial t}(t,v) = \int_{\mathbb{R}^3}\int_{S^2} \left[f(t,v')f(t,w') - f(t,v)f(t,w)\right] B(v,w,\sigma)\, d\sigma\, dw := Q(f)(t,v)$$

où v' et w' sont les vitesses des particules juste avant une collision donnant des particules de vitesse v et w et où la fonction B est appelée noyau de collision et quantifie la probabilité d'obtenir un paramètre $\sigma \in S^2$ dans la collision impliquant deux particules de vitesse v et w. Il s'agit d'une fonction positive.

Nous allons faire des hypothèses simplificatrices sur le noyau de collision B. Par des arguments physiques de symétrie, le noyau de collision B ne doit dépendre que de la vitesse relative $|v-w|$ et du cosinus de l'angle de déviation :

$$cos(\theta) = \langle k, \sigma \rangle$$

où $k := \frac{v-w}{|v-w|}$ et $\langle \, , \, \rangle$ est le produit scalaire usuel sur $\mathbb{R}^3$. Nous supposons que le noyau de collision dépend en fait seulement du cosinus de l'angle de déviation. Un tel noyau $B(cos(\theta))$ sera appelé noyau maxwellien.

On supposera enfin le noyau B intégrable dans le sens où :

$$\forall k \in S^2, \quad \int_{S^2} B\left(\langle k, \sigma \rangle\right) \, d\sigma = 2\pi \int_0^\pi B(cos(\theta)) sin(\theta) \, d\theta < +\infty.$$

Cette hypothèse est appelée condition de cut-off. Pour simplifier, on supposera que cette quantité est égale à 4π, quitte à renormaliser B.

Enfin, on obtient une formulation faible de l'équation de Boltzmann en intégrant f contre $\phi \in \mathbb{C}_c^0(\mathbb{R}^3)$, et en utilisant le changement de variable dit pré-postcollisionnel, qui est involutif et de jacobien 1, comme on peut le voir dans [6] : $(v, w, \sigma) \mapsto (v', w', k)$:

$$\begin{aligned} \int_{\mathbb{R}^3} \phi(v) f(t, v) \, dv = & \int_{\mathbb{R}^3} \phi(v) f(0, v) \, dv \\ & + \int_0^t \int_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} (\phi(v') - \phi(v)) f(s, v) f(s, w) B(v, w, \sigma) \, \frac{dv \, dw \, d\sigma \, ds}{4\pi}. \end{aligned}$$

Cette formulation conduit naturellement à la définition suivante.

Définition 3.1.
On se donne $f \in \mathcal{P}(\mathbb{R}^3)$ et on définit les deux mesures suivantes par dualité grâce au théorème 4.1 de dualité de Riesz (voir l'annexe à la fin).

- $\forall \phi \in C_c(\mathbb{R}^3), \quad \langle Q(f), \phi \rangle = \displaystyle\int_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} [\phi(v') - \phi(v)] \, B(cos(\theta)) \, df(v) \, df(w) \, \frac{d\sigma}{4\pi}$
- $\forall \phi \in C_c(\mathbb{R}^3), \quad \langle Q_+(f), \phi \rangle = \displaystyle\int_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} \phi(v') \, B(cos(\theta)) \, df(v) \, df(w) \, \frac{d\sigma}{4\pi}$

où on note toujours $v' = \frac{v+w}{2} + \frac{|v-w|}{2}\sigma$.

Remarque 3.2. - *Q_+ est une probabilité. En effet, d'après la fin du théorème 4.1 de dualité de Riesz, en appliquant la formule à une fonction continue à support compact, à valeurs dans $[0, 1]$ et valant 1 sur la boule $B(0, R)$, on a :*

$$\forall R > 0, \qquad \int_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} \mathbf{1}_{B(0,R)}(v') \, B(cos(\theta)) \, df(v) \, df(w) \, \frac{d\sigma}{4\pi} \leq Q_+(f)(\mathbb{R}^3) \leq 1,$$

la dernière inégalité étant obtenue car :

$$\int_{S^2} B \, \frac{d\sigma}{4\pi} = 1.$$

En laissant tendre R vers $+\infty$, on obtient le résultat voulu par convergence monotone (puisque B est positive).

- *On a :* $Q(f) = Q_+(f) - f$. *En effet, si* $\phi \in C_c(\mathbb{R}^3)$, *on a :*

$$\begin{aligned}
\langle Q_+(f) - f, \phi \rangle &= \int\limits_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} \phi(v')\, B(cos(\theta))\, df(v)\, df(w)\, \frac{d\sigma}{4\pi} - \int_{\mathbb{R}^3} \phi(v)\, df(v) \\
&= \int\limits_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} [\phi(v') - \phi(v)]\, B(cos(\theta))\, df(v)\, df(w)\, \frac{d\sigma}{4\pi} \\
&= \langle Q(f), \phi \rangle.
\end{aligned}$$

Ainsi l'équation de Boltzmann (dans sa formulation faible) s'écrit :

$$\frac{\partial f}{\partial t} = Q(f) = Q_+(f) - f \quad (B),$$

au sens faible des probabilités.

Démontrons un lemme clé qui sera utile dans la suite :

Lemme 3.3. *Soient* $f, g \in \mathcal{P}_2(\mathbb{R}^3)$, *alors on a :*

$$W_2(Q_+(f), Q_+(g)) \leq W_2(f, g).$$

Démonstration. Pour simplifier la preuve, on se place uniquement dans le cas où B est constant égal à 1. Soit $\phi \in C_c(\mathbb{R}^3)$, par définition on a :

$$\langle Q_+(f), \phi \rangle = \int\limits_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} \phi(v')\, df(v)\, df(w)\, \frac{d\sigma}{4\pi}.$$

Ainsi lorsqu'on fixe v et w, et que σ parcourt uniformément S^2 , v' parcourt uniformément la sphère $S(v, w)$ de centre $\frac{v+w}{2}$ et de rayon $\frac{|v-w|}{2}$. On a alors, en notant $\mathcal{U}(v, w)$ la probabilité uniforme sur $S(v, w)$:

$$Q_+(f) = \int\limits_{\mathbb{R}^3 \times \mathbb{R}^3} \mathcal{U}(v, w)\, df(v)\, df(w)$$

au sens où

$$\forall \phi \in C_c(\mathbb{R}^3), \quad \langle Q_+(f), \phi \rangle = \int\limits_{\mathbb{R}^3 \times \mathbb{R}^3} \langle \mathcal{U}(v, w), \phi \rangle \, df(v)\, df(w)$$

comme défini en annexe.

D'après le théorème 1.13, il existe un plan de transport optimal $\pi \in \mathcal{P}(\mathbb{R}^6)$ transportant g sur f. On peut alors réécrire :

$$\left\{ \begin{array}{ll} Q_+(f) &= \displaystyle\int_{\mathbb{R}^{12}} \mathcal{U}(v, w)\, d\pi(x, v)\, d\pi(y, w) \\[2ex] Q_+(g) &= \displaystyle\int_{\mathbb{R}^{12}} \mathcal{U}(x, y)\, d\pi(x, v)\, d\pi(y, w). \end{array} \right. \qquad (*)$$

Ainsi, par le théorème 2.16 de convexité de W_2^2, on a :

$$\begin{aligned} W_2^2(Q_+(f), Q_+(g)) &= W_2^2\left(\int_{\mathbb{R}^{12}} \mathcal{U}(v,w)\, d\pi(x,v)\, d\pi(y,w), \int_{\mathbb{R}^{12}} \mathcal{U}(x,y)\, d\pi(x,v)\, d\pi(y,w)\right) \\ &\leq \int_{\mathbb{R}^{12}} W_2^2(\mathcal{U}(v,w), \mathcal{U}(x,y))\, d\pi(x,v)\, d\pi(y,w). \end{aligned}$$

Il nous reste à majorer la quantité $W_2^2(\mathcal{U}(v,w), \mathcal{U}(x,y))$ pour chaque v, w, x, y. Pour simplifier les notations, on va majorer le coût de transport de la probabilité uniforme $\mathcal{U}_S$ sur la sphère S de centre O et de rayon r sur la probabilité uniforme $\mathcal{U}_{S'}$ sur la sphère S' de centre O' et de rayon r'.

On vérifie que l'homothétie T de centre $\Omega = O + \frac{r}{r'-r}(O'-O)$ et de rapport $\frac{r'}{r}$ définie par :

$$\forall v, \quad T(v) = \frac{r'}{r}(v - \Omega) + \Omega$$

envoie S sur S'. De plus, on a : $\mathcal{U}_{S'} = T * \mathcal{U}_S$. Alors :

$$\begin{aligned} W_2^2(\mathcal{U}_S, \mathcal{U}_{S'}) &\leq \int_S |T(v) - v|^2\, d\mathcal{U}_S(v) \\ &= \int_S \left| \left(\frac{r'}{r} - 1\right)(v - \Omega) \right|^2 d\mathcal{U}_S(v) \\ &= \left(\frac{r'}{r} - 1\right)^2 \int_S |v - \Omega|^2 d\mathcal{U}_S(v) \\ &= \left(\frac{r'}{r} - 1\right)^2 \int_{S^2} \left| rx - \frac{r}{r'-r}(O' - O) \right|^2 d\mathcal{U}_{S^2}(x) \\ &= |r' - r|^2 + |O' - O|^2. \end{aligned}$$

La dernière égalité est obtenue en développant la norme au carré et en remarquant que la loi uniforme sur la sphère unité S^2 est d'espérance nulle. En appliquant cette majoration à notre problème de transport, on déduit :

$$\begin{aligned} W_2^2(\mathcal{U}(v,w), \mathcal{U}(x,y)) &\leq \left| \frac{v+w}{2} - \frac{x+y}{2} \right|^2 + \left| \frac{|v-w|}{2} - \frac{|x-y|}{2} \right|^2 \\ &\leq \frac{1}{4}\left(|v - x + (w-y)|^2 + |v - w - (x-y)|^2\right) \text{ (Cauchy-Schwarz)} \\ &= \frac{1}{4}\left(|v - x + (w-y)|^2 + |v - x - (w-y)|^2\right) \\ &= \frac{1}{2}(|v-x|^2 + |w-y|^2). \end{aligned}$$

Cette inégalité montre la continuité des applications $(v,w,x,y) \mapsto \mathcal{U}(v,w) \in \mathcal{P}_2(\mathbb{R}^3)$ et $(v,w,x,y) \mapsto \mathcal{U}(x,y) \in \mathcal{P}_2(\mathbb{R}^3)$, donc leur mesurabilité. Ceci justifie que les écritures au sens faible $(*)$ ont bien un sens.

Ainsi en reprenant l'inégalité de convexité du théorème 2.16, on a :

$$\begin{aligned}
W_2^2(Q_+(f), Q_+(g)) &\leq \frac{1}{2} \int_{\mathbb{R}^{12}} (|v-x|^2 + |w-y|^2)\, d\pi(x,v)\, d\pi(y,w) \\
&= \int_{\mathbb{R}^6} |v-x|^2\, d\pi(x,v) \\
&= W_2^2(f,g).
\end{aligned}$$

□

3.2 Théorème d'existence et d'unicité

Dans un premier temps, on se donne $T > 0$ et on montre l'existence et l'unicité des solutions à l'équation de Boltzmann sur $[0,T]$. A la fin, on aura l'existence et l'unicité sur $\mathbb{R}^+$ grâce à l'unicité des solutions sur un intervalle de temps borné démontrée précédemment.

Définition 3.4. *Posons d'abord :*

$$C_T := C([0,T], \mathcal{P}_2(\mathbb{R}^3)).$$

On se donne $f_0 \in \mathcal{P}_2(\mathbb{R}^3)$. On dit que $f \in C_T$ est solution de (B) sur $[0,T]$, de donnée initiale f_0, si :

$$\forall t \in [0,T], \quad f(t) = f_0\, e^{-t} + \int_0^t e^{s-t}\, Q_+(f(s))\, ds \quad (1)$$

au sens faible des probabilités, c'est-à-dire qu'on a :

$$\forall \phi \in C_c(\mathbb{R}^3), \quad \forall t \in [0,T], \quad \langle f(t), \phi \rangle = e^{-t} \langle f_0, \phi \rangle + \int_0^t e^{s-t}\, \langle Q_+(f(s)), \phi \rangle\, ds.$$

Ou de manière équivalente (on ne le montrera pas) si :

$$\forall t \in [0,T], \quad f(t) = f_0 + \int_0^t Q(f(s))\, ds \quad (2).$$

Cette formulation correspond à la formulation faible de l'équation de Boltzmann formulée au début de cette partie.

L'espace C_T est muni de la distance :

$$\forall f, g \in C_T, \quad d(f,g) = \sup_{0 \leq t \leq T} W_2(f(t), g(t)).$$

Puisque $(\mathcal{P}_2(\mathbb{R}^3), W_2)$ est complet, (C_T, d) est également complet.

Montrons tout d'abord que la définition des solutions a bien un sens.

Lemme 3.5. *Soit $f \in C_T$, alors l'application*

$$\left\{ \begin{array}{ccc} [0,T] & \to & \mathcal{P}_2(\mathbb{R}^3) \\ t & \mapsto & Q_+(f(t)) \end{array} \right.$$

appartient à C_T. De plus on a :

$$\int_{\mathbb{R}^3} |v|^2\, dQ_+(\mu)(v) = \int_{\mathbb{R}^3} |v|^2\, d\mu(v).$$

Démonstration. Démontrons tout d'abord que :

$$\int_{\mathbb{R}^3\times\mathbb{R}^3\times S^2} B\left(\left\langle \frac{v-w}{|v-w|},\sigma\right\rangle\right)|v-w|\langle v+w,\sigma\rangle\, df(v)\, df(w)\,\frac{d\sigma}{4\pi}=0,$$

ce qui sera utile dans le calcul qui suit. En utilisant le théorème de Fubini, on réécrit cette quantité :

$$\int_{\mathbb{R}^6}|v-w|\left\langle (v+w),\int_{S^2}\sigma B\left(\left\langle \frac{v-w}{|v-w|},\sigma\right\rangle\right)\frac{d\sigma}{4\pi}\right\rangle df(v)\, df(w).$$

On se place en coordonnées sphériques d'axe $\vec{e_z} := \frac{v-w}{|v-w|}$. On obtient :

$$\begin{aligned}
&\int_{S^2}\sigma B\left(\left\langle \frac{v-w}{|v-w|},\sigma\right\rangle\right)\frac{d\sigma}{4\pi}\\
&=\int_0^{\pi}\int_0^{2\pi} B(cos(\theta))sin(\theta)\vec{e_r}\, d\phi\, d\theta\\
&=\int_0^{\pi}\int_0^{2\pi} B(cos(\theta))sin(\theta)(cos(\theta)\vec{e_z}+sin(\theta)cos(\phi)\vec{e_x}+sin(\theta)sin(\phi)\vec{e_y})\, d\phi\, d\theta\\
&=\int_0^{\pi} B(cos(\theta))sin(\theta)cos(\theta)\,\frac{d\theta}{2}\,\vec{e_z}\\
&:= C_b\,\vec{e_z}.
\end{aligned}$$

Ainsi :

$$\begin{aligned}
&\int_{\mathbb{R}^6}|v-w|\left\langle (v+w),\int_{S^2}\sigma B\left(\left\langle \frac{v-w}{|v-w|},\sigma\right\rangle\right)\frac{d\sigma}{4\pi}\right\rangle df(v)\, df(w)\\
&= C_b\int_{\mathbb{R}^6}\langle v+w, v-w\rangle\, df(v)\, df(w)\\
&= C_b\int_{\mathbb{R}^6}(|v|^2-|w|^2)\, df(v)\, df(w)\\
&=0.
\end{aligned}$$

D'où le résultat.

1. Soit $\mu\in\mathcal{P}_2(\mathbb{R}^3)$. Montrons que $Q_+(\mu)\in\mathcal{P}_2(\mathbb{R}^3)$. On a déjà vu à la remarque 3.2 que

$Q_+(\mu)$ est une probabilité. De plus, on a :

$$\begin{aligned}
\int_{\mathbb{R}^3} |v|^2 \, dQ_+(\mu)(v) &= \int\limits_{\mathbb{R}^3\times\mathbb{R}^3\times S^2} |v'|^2 \, B(cos(\theta)) \, df(v) \, df(w) \, \frac{d\sigma}{4\pi} \\
&= \int\limits_{\mathbb{R}^3\times\mathbb{R}^3\times S^2} \left| \frac{v+w}{2} + \frac{|v-w|}{2}\sigma \right|^2 B(cos(\theta)) \, df(v) \, df(w) \, \frac{d\sigma}{4\pi} \\
&= \frac{1}{4} \int\limits_{\mathbb{R}^3\times\mathbb{R}^3\times S^2} (|v+w|^2 + |v-w|^2) \, df(v) \, df(w) \, \frac{d\sigma}{4\pi} \\
&\quad + \frac{1}{2} \int_{\mathbb{R}^3\times\mathbb{R}^3} \left\langle |v-w|(v+w) \, , \, \left(\int_{S^2} \sigma B(cos(\theta)) \frac{d\sigma}{4\pi} \right) \right\rangle \, df(v) \, df(w) \\
&= \frac{1}{4} \int\limits_{\mathbb{R}^3\times\mathbb{R}^3\times S^2} 2(|v|^2 + |w|^2) \, df(v) \, df(w) \, \frac{d\sigma}{4\pi} \\
&= \int_{\mathbb{R}^3} |v|^2 \, d\mu(v).
\end{aligned}$$

Ainsi $Q_+(\mu) \in \mathcal{P}_2(\mathbb{R}^3)$.

2. Il reste à montrer la continuité. Soit $t_0 \in [0,T]$ et $t_n \to t_0$ une suite de $[0,T]$. D'après le lemme 3.3, on a :

$$\forall n \in \mathbb{N}, \quad W_2(Q_+(f(t_n)), Q_+(f(t_0))) \leq W_2(f(t_n), f(t_0)) \quad \underset{n\to+\infty}{\to} 0,$$

par continuité de f.

□

Définissons l'opérateur qui va permettre de montrer l'existence de solutions à l'équation de Boltzmann. Les solutions seront obtenues comme points fixes de l'application suivante.

Définition 3.6. *On définit l'opérateur T par :*

$$T : \left\{ \begin{array}{lll} C_T & \to & C_T \\ f & \mapsto & T(f) : t \mapsto e^{-t} \, f(0) + \displaystyle\int_0^t e^{s-t} \, Q_+(f(s)) \, ds. \end{array} \right.$$

Montrons que T est bien à valeurs dans C_T.

Démonstration.

- Soient $f \in C_T$ et $t \in [0,T]$. Comme $f_0 \in \mathcal{P}_2(\mathbb{R}^3)$ et que d'après le lemme 3.5, $Q_+(f(s)) \in \mathcal{P}_2(\mathbb{R}^3)$, on en déduit que $T(f)(t) \in \mathcal{P}_2(\mathbb{R}^3)$ comme combinaison convexe d'éléments de $\mathcal{P}_2(\mathbb{R}^3)$.
- Il reste à montrer la continuité de $T(f)$. Pour cela, on fixe $t_0 \in [0,T]$ et on considère une suite de $[0,T]$ telle que $t_n \to t_0$. Il suffit de montrer, d'après le théorème 2.10 :

(a) $\displaystyle\int_{\mathbb{R}^3} \phi(v) \, dT(f)(t_n)(v) \underset{n\to+\infty}{\to} \int_{\mathbb{R}^3} \phi(v) \, dT(f)(t_0)(v) \quad \forall \phi \in C_c(\mathbb{R}^3)$

(b) $\displaystyle\int_{\mathbb{R}^3} |v|^2 \, dT(f)(t_n)(v) \underset{n\to+\infty}{\rightarrow} \int_{\mathbb{R}^3} |v|^2 \, dT(f)(t_0)(v).$

Pour (a), on prend $\phi \in C_c(\mathbb{R}^3)$ et on écrit pour tout $n \in \mathbb{N}$:

$$\begin{aligned}
&\left| \int_{\mathbb{R}^3} \phi \, dT(f)(t_n) - \int_{\mathbb{R}^3} \phi \, dT(f)(t_0) \right| \\
&\leq \left| \int_{\mathbb{R}^3} \phi \, df_0 \right| |e^{-t_n} - e^{-t_0}| \\
&\quad + \left| \int_0^{t_n} \left(\int_{\mathbb{R}^3} \phi \, dQ_+(f(s)) \right) e^{s-t_n} \, ds - \int_0^{t_0} \left(\int_{\mathbb{R}^3} \phi \, dQ_+(f(s)) \right) e^{s-t_n} \, ds \right| \\
&\quad + \left| \int_0^{t_0} \left(\int_{\mathbb{R}^3} \phi \, dQ_+(f(s)) \right) e^{s-t_n} \, ds - \int_0^{t_0} \left(\int_{\mathbb{R}^3} \phi \, dQ_+(f(s)) \right) e^{s-t_0} \, ds \right| \\
&\leq \sup \left(\left| \int_{\mathbb{R}^3} \phi \, df_0 \right|, \sup_{0 \leq s \leq t_n} \left| \int_{\mathbb{R}^3} \phi \, dQ_+(f(s)) \right| \right) \\
&\quad \times \left(|e^{-t_n} - e^{-t_0}| + |1 - e^{t_0 - t_n}| + (e^{t_0} - 1) |e^{-t_n} - e^{-t_0}| \right) \\
&\leq \|\phi\|_\infty \left(|e^{-t_n} - e^{-t_0}| + |1 - e^{t_0 - t_n}| + (e^{t_0} - 1) |e^{-t_n} - e^{-t_0}| \right) \quad \underset{n\to+\infty}{\rightarrow} 0.
\end{aligned}$$

Pour (b), on fait le même raisonnement et on utilise cette fois la majoration suivante, obtenue grâce au lemme 3.5 et valable pour tout $n \in \mathbb{N}$:

$$\begin{aligned}
\sup \left(\int_{\mathbb{R}^3} |v|^2 \, df_0(v) \, , \sup_{0 \leq s \leq t_n} \int_{\mathbb{R}^3} |v|^2 \, dQ_+(f(s))(v) \right) &= \sup \left(\int_{\mathbb{R}^3} |v|^2 \, df_0(v), \sup_{0 \leq s \leq t_n} \int_{\mathbb{R}^3} |v|^2 \, df(s)(v) \right) \\
&\leq \sup_{0 \leq s \leq T} \int_{\mathbb{R}^3} |v|^2 \, df(s)(v) < +\infty.
\end{aligned}$$

En effet, l'application $t \mapsto \displaystyle\int_{\mathbb{R}^3} |v|^2 \, df(t)(v)$ est continue puisque $f \in C_T$ et en utilisant le théorème 2.10 de convergence en distance de Wasserstein.

D'où $T(f) \in C_T$.

□

A ce stade on a montré que la définition des solutions à l'équation de Boltzmann a bien un sens. On peut enfin démonter l'existence et l'unicité des solutions. En effet il y a équivalence entre être solution de (B) et être un point fixe de T.

Théorème 3.7 (Existence et unicité des solutions)**.**
Soit $f_0 \in \mathcal{P}_2(\mathbb{R}^3)$, alors il existe une unique solution à (B) sur $[0, T]$ (au sens donné par la définition 3.4) de donnée initiale f_0.

De plus, les solutions conservent la quantité de mouvement et l'énergie cinétique :

$$\begin{aligned}
\int_{\mathbb{R}^3} v \, df(t)(v) &= \int_{\mathbb{R}^3} v \, df_0(v), \quad \forall t \in [0, T], \\
\int_{\mathbb{R}^3} |v|^2 df(t)(v) &= \int_{\mathbb{R}^3} |v|^2 df_0(v), \quad \forall t \in [0, T].
\end{aligned}$$

Les conservations sont assez naturelles puisqu'elles sont vérifiées pour chaque particule du gaz (chocs élastiques).

Démonstration.

On démontre seulement l'existence et l'unicité.

Étape 1 : Complétude

On considère l'espace $C_T^{f_0} := \{f \in C_T,\ f(0) = f_0\}$. Cet espace est fermé dans C_T. En effet, soit $(f_n)_n$ une suite de $C_T^{f_0}$ qui converge vers $f \in C_T$. Par définition de la distance sur C_T, on a :

$$f_n(0) = f_0 \underset{W_2}{\rightarrow} f(0).$$

Comme la convergence en distance W_2 implique la convergence étroite (cf théorème 2.10), on en déduit que $f(0) = f_0$. D'où le caractère fermé de l'espace. Ainsi, puisque C_T est complet, on en déduit que $C_T^{f_0}$ est également complet.

Étape 2 : T est une application contractante

Soient $f, g \in C_T^{f_0}$. On a déjà $T(f), T(g) \in C_T^{f_0}$ puisque :

$$T(f)(0) = f(0) = f_0 = g_0.$$

Par convexité de W_2^2, on a pour tout $t \in [0, T]$:

$$\begin{aligned} W_2^2(T(f)(t), T(g)(t)) &\leq e^{-t}\, W_2^2(T(f)(0), T(g)(0)) + \int_0^t e^{s-t}\, W_2^2(Q_+(f(s)), Q_+(g(s)))\, ds \\ &\leq \int_0^t e^{s-t}\, W_2^2(f(s), g(s))\, ds \qquad \text{car } T(f)(0) = T(g)(0) \text{ et d'après le lemme 3.3} \\ &\leq (1 - e^{-T})\, d^2(f, g). \end{aligned}$$

En passant au sup, on obtient la propriété de contraction :

$$d(T(f), T(g)) \leq \sqrt{1 - e^{-T}}\, d(f, g).$$

Étape 3 : Conclusion

On applique le théorème du point fixe de Banach : l'application T admet un unique point fixe qui est donc, par définition, l'unique solution de (B) de donnée initiale f_0.

□

Remarque 3.8. *Soit $n \in \mathbb{N}^*$. Nous avons montré l'existence et l'unicité de la solution f^n de (B) de donnée initiale $f_0 \in \mathcal{P}_2(\mathbb{R}^3)$ sur $[0, n]$. L'unicité assure alors que :*

$$\forall n \geq m, \quad f^n_{|[0,m]} = f^m.$$

Cela assure l'existence et l'unicité de la solution de donnée initiale f_0 sur $[0, +\infty)$.

3.3 Convergence vers l'équilibre et stabilité

Dans la partie précédente, on a montré l'unicité de la solution de donnée initiale f_0 à l'équation de Boltzmann. Le théorème suivant précise cette unicité en donnant un contrôle entre deux solutions.

Théorème 3.9. *Soient $f_0, g_0 \in \mathcal{P}_2(\mathbb{R}^3)$, alors les solutions correspondantes de (B), qu'on note $f(t)$ et $g(t)$, vérifient :*

$$\forall t, \quad W_2(f(t), g(t)) \leq W_2(f_0, g_0)$$

ce qui est équivalent à la décroissance de $t \mapsto W_2(f(t), g(t))$.

Démonstration. On connait l'expression des solutions de (B) :

$$\forall t, \quad f(t) = T(f)(t) = e^{-t} f_0 + \int_0^t e^{s-t} Q_+(f(s)) \, ds.$$

Ainsi par convexité de W_2^2 :

$$\begin{aligned} W_2^2(f(t), g(t)) &\leq e^{-t} W_2^2(f_0, g_0) + \int_0^t W_2^2(Q_+(f(s)), Q_+(g(s))) \, e^{s-t} \, ds \\ &\leq e^{-t} W_2^2(f_0, g_0) + \int_0^t W_2^2(f(s), g(s)) \, e^{s-t} \, ds \quad \text{d'après le lemme 3.3.} \end{aligned}$$

L'application $t \mapsto y(t) := W_2^2(f(t), g(t)) e^t$ est continue (car f est continue). De plus, on vient de montrer que y vérifie :

$$\forall t, \quad y(t) \leq y(0) + \int_0^t y(s) \, ds.$$

Le lemme de Gronwall assure alors que :

$$\forall t \geq 0, \quad W_2(f(t), g(t)) \leq W_2(f_0, g_0).$$

□

Nous avons donc montré la décroissance de la distance de Wasserstein entre deux solutions quelconques. Nous allons nous intéresser au cas d'égalité dans le lemme 3.3, ce qui sera utile dans le théorème de convergence vers l'équilibre qui suit pour identifier la loi des vitesses à l'équilibre.

Lemme 3.10. *Soient $f, g \in \mathcal{P}_2(\mathbb{R}^3)$ vérifiant :*

$$\begin{aligned} \int_{\mathbb{R}^3} v \, df(v) &= \int_{\mathbb{R}^3} v \, dg(v) \\ \int_{\mathbb{R}^3} |v|^2 \, df(v) &= \int_{\mathbb{R}^3} |v|^2 \, dg(v). \end{aligned}$$

On suppose de plus que g est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^3$ notée λ et que $\frac{dg}{d\lambda} > 0 \quad \lambda$-pp. Alors :

$$W_2(Q_+(f), Q_+(g)) = W_2(f, g) \Leftrightarrow f = g.$$

Démonstration. ($\Leftarrow$) Immédiat.

($\Rightarrow$) On note $\pi \in \mathcal{P}(\mathbb{R}^3 \times \mathbb{R}^3)$ un plan de transport optimal envoyant g sur f. On reprend la preuve du lemme 3.3 : on doit avoir des égalités au lieu des inégalités, en particulier lorsqu'on utilise Cauchy-Schwarz :

$$\frac{v-w}{|v-w|} = \frac{x-y}{|x-y|} \quad \pi(x,v) \otimes \pi(y,w)\text{-pp.}$$

Comme g est absolument continue par rapport à λ, le théorème 1.18 de Brenier assure que le plan de transport optimal π est unique et induit par une application de transport $u : \mathbb{R}^3 \to \mathbb{R}^3$. En particulier, $f = u * g$ et $d\pi(x,v) = \delta_{u(x)}(v) \otimes dg(x)$. Ainsi :

$$\frac{x-y}{|x-y|} = \frac{u(x)-u(y)}{|u(x)-u(y)|} \quad g(x) \otimes g(y)\text{-pp} \quad (*).$$

Cette égalité est également vraie $\lambda \otimes \lambda$-pp puisque $g << \lambda$ et $\frac{dg}{d\lambda} > 0 \quad \lambda$-pp. On considère alors un $x_0 \in \mathbb{R}^3$ tel que :

$$u(x) = u(x_0) + \frac{|u(x)-u(x_0)|}{|x-x_0|}\,(x-x_0) \quad \lambda(x)\text{-pp} \quad (**).$$

Alors, pour λ presque tous $x_1, x_2 \in \mathbb{R}^3$, on a :

$$\begin{aligned}\frac{|u(x_1)-u(x_0)|}{|x_1-x_0|}\,(x_1-x_0) - \frac{|u(x_2)-u(x_0)|}{|x_2-x_0|}\,(x_2-x_0) &= u(x_1)-u(x_2)\\ &= \frac{|u(x_1)-u(x_2)|}{|x_1-x_2|}\,(x_1-x_2) \quad \text{d'aprèes } (*).\end{aligned}$$

Puisque (x_0, x_1, x_2) est libre dans $\mathbb{R}^3$ pour λ-presque-tous x_1 et x_2, le coefficient devant x_0 doit être nul, c'est-à-dire :

$$\frac{|u(x_1)-u(x_0)|}{|x_1-x_0|} = \frac{|u(x_2)-u(x_0)|}{|x_2-x_0|} \quad \lambda(x_1) \otimes \lambda(x_2)\text{-pp.}$$

La quantité $\frac{|u(x)-u(x_0)|}{|x-x_0|}$ est donc constante $\lambda(x)$-pp. En reprenant $(**)$, il existe $\alpha \in \mathbb{R}^3$ et $\beta \in \mathbb{R}$ tels que :

$$u(x) = \alpha + \beta x \quad \lambda(x)\text{-pp.}$$

Par un calcul, les hypothèses sur les deux premiers moments de f et g assurent alors que $\alpha = 0$ et $\beta^2 = 1$. En reprenant la condition $(*)$, on trouve $\beta = 1$. Ainsi $u = Id_{\mathbb{R}^3} \quad \lambda$-pp et donc g-pp. Finalement, on trouve $f = g$ puisque $f = u * g$.

□

Nous allons maintenant nous intéresser à la convergence en temps grand des solutions de l'équation de Boltzmann vers un état d'équilibre.

Définition 3.11. *Un équilibre de l'équation de Boltzmann est une probabilité $\mu \in \mathcal{P}_2(\mathbb{R}^3)$ telle que, si $f_0 = \mu$, alors pour tout $t > 0$, $f(t) = \mu$.*

Donnons un exemple d'équilibres de l'équation de Boltzmann, ce qu'on peut montrer par un calcul explicite.

Proposition 3.12. *Soit G une gaussienne de moyenne $m \in \mathbb{R}^3$ et de covariance $\sigma^2 Id_{\mathbb{R}^3}$, avec $\sigma > 0$ (G est appelée maxwellienne). Alors G est un équilibre de (B).*

Dans la suite, on se restreint au cas où le noyau de collision est constant, c'est-à-dire $B = 1$. Démontrons, avant de traiter le théorème de convergence vers l'équilibre, que si la donnée initiale à un moment d'ordre 4 fini, alors les moments d'ordre 4 de la solution de (B) sont uniformément bornés en temps.

Lemme 3.13. *On se place dans le cas où $B = 1$. Soit $f_0 \in \mathcal{P}_4(\mathbb{R}^3)$, alors :*

$$\sup_{t\geq 0} \int_{\mathbb{R}^3} |v|^4 \, df(t)(v) < +\infty.$$

Etape 1 : On montre que :

$$\forall f \in \mathcal{P}(\mathbb{R}^3), \quad \int_{\mathbb{R}^3} |v|^4 \, dQ_+(f)(v) \leq \frac{2}{3} \int_{\mathbb{R}^3} |v|^4 \, df(v) + 3 \left(\int_{\mathbb{R}^3} |v|^2 \, df(v) \right)^2.$$

Par définition, on a :

$$\int_{\mathbb{R}^3} |v|^4 \, dQ_+(f)(v) = \int_{\mathbb{R}^6 \times S^2} \left| \frac{v+w}{2} + \frac{|v-w|}{2}\sigma \right|^4 df(v) \, df(w) \, \frac{d\sigma}{4\pi}.$$

Or en développant d'abord un carré avec la formule usuelle et en élevant le résultat au carré, on a :

$$\begin{aligned} \left| \frac{v+w}{2} + \frac{|v-w|}{2}\sigma \right|^4 &= \left| \frac{v+w}{2} \right|^4 + \left| \frac{v-w}{2} \right|^4 && (1) \\ &\quad + \left| \frac{v-w}{2} \right|^2 \langle v+w, \sigma \rangle^2 && (2) \\ &\quad + 2 \left| \frac{v+w}{2} \right|^2 \left| \frac{v-w}{2} \right|^2 && (3) \\ &\quad + \left| \frac{v+w}{2} \right|^2 |v-w| \langle v+w, \sigma \rangle && (4) \\ &\quad + \left| \frac{v-w}{2} \right|^2 |v-w| \langle v+w, \sigma \rangle. && (5) \end{aligned}$$

Étudions chacun des cinq termes successivement.

1.

$$\begin{aligned} \int (1) &= \frac{1}{16} \int_{\mathbb{R}^6} |v+w|^4 + |v-w|^4 \, df(v) \, df(w) \\ &= \frac{1}{16} \int_{\mathbb{R}^6} (|v|^2 + |w|^2 + 2\langle v, w \rangle)^2 + (|v|^2 + |w|^2 - 2\langle v, w \rangle)^2 \, df(v) \, df(w) \\ &= \frac{1}{16} \int_{\mathbb{R}^6} 2|v|^4 + 2|w|^4 + 8\langle v, w \rangle^2 + 4|v|^2|w|^2 \, df(v) \, df(w) \\ &\leq \frac{1}{16} \int_{\mathbb{R}^6} 2|v|^4 + 2|w|^4 + 12|v|^2|w|^2 \, df(v) \, df(w) \quad \text{(Cauchy-Schwarz)} \\ &\leq \frac{1}{4} \int_{\mathbb{R}^3} |v|^4 \, df(v) + \left(\int_{\mathbb{R}^3} |v|^2 \, df(v) \right)^2. \end{aligned}$$

2. Pour ce terme, on sépare l'intégrale sur la sphère grâce à Fubini et on se place en coordonnées sphériques d'axe vertical $\frac{v+w}{|v+w|}$.

$$\begin{aligned}
\int (2) &= \int_{\mathbb{R}^6} \left| \frac{v-w}{2} \right|^2 \left(\int_{S^2} \langle v+w, \sigma \rangle^2 \frac{d\sigma}{4\pi} \right) df(v)\, df(w) \\
&= \int_{\mathbb{R}^6} \left| \frac{v-w}{2} \right|^2 |v+w|^2 \left(\int_0^{2\pi} \int_0^{\pi} (cos(\theta))^2 sin(\theta)\, d\theta\, \frac{d\phi}{4\pi} \right) df(v)\, df(w) \\
&= \int_{\mathbb{R}^6} \left| \frac{v-w}{2} \right|^2 \frac{|v+w|^2}{3}\, df(v)\, df(w) \\
&= \frac{1}{12} \int_{\mathbb{R}^6} (|v|^2 + |w|^2 - 2\langle v, w\rangle)(|v|^2 + |w|^2 + 2\langle v, w\rangle)\, df(v)\, df(w) \\
&= \frac{1}{12} \int_{\mathbb{R}^6} |v|^4 + |w|^4 + 2|v|^2|w|^2 - 4\langle v, w\rangle^2\, df(v)\, df(w) \\
&\leq \frac{1}{6} \int_{\mathbb{R}^3} |v|^4\, df(v) + \left(\int_{\mathbb{R}^3} |v|^2\, df(v) \right)^2 .
\end{aligned}$$

3. On a :

$$\begin{aligned}
\int (3) &= \frac{1}{8} \int_{\mathbb{R}^6} |v+w|^2 |v-w|^2\, df(v)\, df(w) \\
&= \frac{1}{8} \int_{\mathbb{R}^6} |v|^4 + |w|^4 + 2|v|^2|w|^2 - 4\langle v, w\rangle^2\, df(v)\, df(w) \quad \text{comme dans le calcul de 2.} \\
&\leq \frac{1}{8} \int_{\mathbb{R}^6} |v|^4 + |w|^4 + 6|v|^2|w|^2\, df(v)\, df(w) \quad \text{(Cauchy-Schwarz)} \\
&\leq \frac{1}{4} \int_{\mathbb{R}^3} |v|^4\, df(v) + \left(\int_{\mathbb{R}^3} |v|^2\, df(v) \right)^2 .
\end{aligned}$$

4. Pour ce terme, on sépare l'intégrale sur la sphère grâce à Fubini et on se place en coordonnées sphériques d'axe vertical $\frac{v+w}{|v+w|}$.

$$\begin{aligned}
\int (4) &= \int_{\mathbb{R}^6} \left| \frac{v+w}{2} \right|^2 |v-w| \left(\int_{S^2} \langle v+w, \sigma \rangle \frac{d\sigma}{4\pi} \right) df(v)\, df(w) \\
&= \int_{\mathbb{R}^6} \left| \frac{v+w}{2} \right|^2 |v-w||v+w| \left(\int_0^{\pi} cos(\theta) sin(\theta) \frac{d\theta}{2} \right) df(v)\, df(w) \\
&= 0.
\end{aligned}$$

5. Le même raisonnement que dans le point précédent montre que :

$$\int (5) = 0.$$

En rassemblant les cinq termes, on trouve finalement que :

$$\int_{\mathbb{R}^3} |v|^4\, dQ_+(f)(v) \leq \frac{2}{3} \int_{\mathbb{R}^3} |v|^4\, df(v) + 3 \left(\int_{\mathbb{R}^3} |v|^2\, df(v) \right)^2 .$$

Étape 2 : On note $\tilde{f}_0$ l'application constante égale à $f_0 \in \mathcal{P}_4(\mathbb{R}^3)$. Le théorème du point fixe de Banach assure que :

$$\forall t \geq 0, \quad T^n(\tilde{f}_0)(t) \rightharpoonup f(t),$$

où $f(t)$ est la solution de (B) à l'instant t. Le lemme 1.15 assure alors :

$$\forall t \geq 0, \quad \int_{\mathbb{R}^3} |v|^4 df(t)(v) \leq \underline{\lim} \int_{\mathbb{R}^3} |v|^4 dT^n(\tilde{f}_0)(t)(v).$$

Pour montrer le résultat il suffit de majorer le terme de droite uniformément en temps. Pour ce faire on a besoin du lemme démontré dans l'étape suivante.

Étape 3 : On montre que pour tout entier n :

$$\forall t \geq 0, \quad \int_{\mathbb{R}^3} |v|^2 \, dT^n(\tilde{f}_0)(t)(v) = \int_{\mathbb{R}^3} |v|^2 \, df_0(v).$$

On raisonne par récurrence, l'initialisation étant immédiate. On suppose le résultat vrai au rang n. Alors pour $t \geq 0$:

$$\begin{aligned}
&\int_{\mathbb{R}^3} |v|^2 \, dT^{n+1}(\tilde{f}_0)(t)(v) \\
&= e^{-t} \int_{\mathbb{R}^3} |v|^2 dT^n(\tilde{f}_0)(0)(v) + \int_0^t e^{s-t} \int_{\mathbb{R}^3} |v|^2 dQ_+(T^n(\tilde{f}_0)(s))(v) \, ds \\
&= e^{-t} \int_{\mathbb{R}^3} |v|^2 dT^n(\tilde{f}_0)(0)(v) + \int_0^t e^{s-t} \int_{\mathbb{R}^3} |v|^2 dT^n(\tilde{f}_0)(s)(v) \, ds \quad \text{d'après le lemme 3.5} \\
&= \int_{\mathbb{R}^3} |v|^2 \, df_0(v).
\end{aligned}$$

La dernière égalité est obtenue car on montre par une récurrence immédiate que pour tout n, $T^n(\tilde{f}_0)(0) = f_0$, et en utilisant l'hypothèse de récurrence.

Étape 4 : Notons $m_0 := \int_{\mathbb{R}^3} |v|^4 \, df_0(v) < +\infty$ et $K := 3 \left(\int_{\mathbb{R}^3} |v|^2 \, df_0(v) \right)^2 < +\infty$. On montre par récurrence que pour tout entier $n \geq 1$, on a :

$$\forall t \geq 0, \quad \int_{\mathbb{R}^3} |v|^4 dT^n(\tilde{f}_0)(t)(v) \leq \sum_{k=1}^{n} \left(\frac{2}{3}\right)^k \frac{t^{k-1}}{(k-1)!} e^{-t} m_0 + \sum_{k=0}^{n} \left(\frac{2}{3}\right)^k (m_0 + K).$$

- Pour $n=1$, on a pour tout $t \geq 0$:

$$\begin{aligned}
\int_{\mathbb{R}^3} |v|^4 dT(\tilde{f}_0)(t)(v) &= e^{-t} m_0 + \int_0^t e^{s-t} \int_{\mathbb{R}^3} |v|^4 dQ_+(\tilde{f}_0(s))(v)\, ds \quad \text{par définition de } T \\
&= e^{-t} m_0 + \int_0^t e^{s-t} \int_{\mathbb{R}^3} |v|^4 dQ_+(f_0)(v)\, ds \\
&\leq e^{-t} m_0 + \int_0^t e^{s-t} \left(\frac{2}{3} m_0 + K \right) ds \quad \text{d'après l'étape 1} \\
&= e^{-t} m_0 + (1-e^{-t}) \left(\frac{2}{3} m_0 + K \right) \\
&\leq \frac{2}{3} m_0 + K + \frac{m_0 e^{-t}}{3} \\
&\leq \frac{2 m_0 e^{-t}}{3} + (m_0 + K) + \frac{2}{3}(m_0 + K).
\end{aligned}$$

Ceci achève l'initialisation.

- On suppose le résultat vrai au rang n (HR). Alors on a pour tout $t \geq 0$:

$$\begin{aligned}
&\int_{\mathbb{R}^3} |v|^4 dT^{n+1}(\tilde{f}_0)(t)(v) \\
&= e^{-t} \int_{\mathbb{R}^3} |v|^4 \, dT^n(\tilde{f}_0)(0)(v) + \int_0^t e^{s-t} \int_{\mathbb{R}^3} |v|^4 \, dQ_+(T^n(\tilde{f}_0)(s))(v)\, ds \\
&\leq e^{-t} m_0 + \int_0^t e^{s-t} \left(\frac{2}{3} \int_{\mathbb{R}^3} |v|^4 \, dT^n(\tilde{f}_0(s))(v) + 3 \left(\int_{\mathbb{R}^3} |v|^2 \, dT^n(\tilde{f}_0)(s)(v) \right)^2 \right) ds \quad \text{(étape 1)} \\
&= e^{-t} m_0 + \int_0^t e^{s-t} \left(\frac{2}{3} \int_{\mathbb{R}^3} |v|^4 \, dT^n(\tilde{f}_0)(s)(v) + K \right) ds \quad \text{d'après l'étape 3} \\
&\leq e^{-t} m_0 + \int_0^t e^{s-t} \left(\frac{2}{3} \left(\sum_{k=1}^{n} \left(\frac{2}{3}\right)^k \frac{s^{k-1}}{(k-1)!} e^{-s} m_0 + \sum_{k=0}^{n} \left(\frac{2}{3}\right)^k (m_0+K) \right) + K \right) ds \quad (HR) \\
&= e^{-t} m_0 + e^{-t} m_0 \sum_{k=1}^{n} \left(\frac{2}{3}\right)^{k+1} \frac{t^k}{k!} + \sum_{k=0}^{n} \left(\frac{2}{3}\right)^{k+1} (m_0+K) + K(1-e^{-t}) \\
&\leq m_0 + K + e^{-t} m_0 \sum_{k=1}^{n} \left(\frac{2}{3}\right)^{k+1} \frac{t^k}{k!} + \sum_{k=0}^{n} \left(\frac{2}{3}\right)^{k+1} (m_0+K) \\
&= e^{-t} m_0 \sum_{k=2}^{n+1} \left(\frac{2}{3}\right)^k \frac{t^{k-1}}{(k-1)!} + (m_0+K) + \sum_{k=1}^{n+1} \left(\frac{2}{3}\right)^k (m_0+K) \\
&\leq e^{-t} m_0 \sum_{k=1}^{n+1} \left(\frac{2}{3}\right)^k \frac{t^{k-1}}{(k-1)!} + \sum_{k=0}^{n+1} \left(\frac{2}{3}\right)^k (m_0+K)
\end{aligned}$$

D'où l'hérédité.

Étape 5 : Conclusion

On a pour tout $t \geq 0$:

$$\begin{aligned}
&\int_{\mathbb{R}^3} |v|^4 df(t)(v) \\
&\leq \underline{\lim} \int_{\mathbb{R}^3} |v|^4 dT^n(\tilde{f}_0)(t)(v) \quad \text{(Etape 2)} \\
&\leq \underline{\lim} \left(\sum_{k=1}^{n} \left(\frac{2}{3}\right)^k \frac{t^{k-1}}{(k-1)!} e^{-t} m_0 + \sum_{k=0}^{n} \left(\frac{2}{3}\right)^k (m_0 + K) \right) \quad \text{(Etape 4)} \\
&= \frac{2}{3} e^{-t/3} m_0 + 3(m_0 + K)
\end{aligned}$$

Cela achève la preuve.

On peut maintenant passer au théorème.

Théorème 3.14. *Soient $f_0, g_0 \in \mathcal{P}_2(\mathbb{R}^3)$ avec*

$$\begin{aligned}
m &:= \int_{\mathbb{R}^3} v \, df_0(v) = \int_{\mathbb{R}^3} v \, dg_0(v) \\
r^2 &:= \int_{\mathbb{R}^3} |v|^2 \, df_0(v) = \int_{\mathbb{R}^3} |v|^2 \, dg_0(v).
\end{aligned}$$

Alors :

$$W_2(f(t), g(t)) \underset{t \to +\infty}{\to} 0.$$

En particulier, si G est la maxwellienne de moyenne m et de covariance $\sigma^2 \, Id_{\mathbb{R}^3}$, où $\sigma^2 = r^2 - |m|^2$, on a :

$$W_2(f(t), G) \underset{t \to +\infty}{\to} 0.$$

Démonstration. On va seulement démontrer ce théorème dans le cas $B = 1$. Par inégalité triangulaire, il suffit de montrer la convergence en distance W_2 des solutions de (B) vers la maxwellienne adaptée.

Étape 1 : On suppose d'abord que $f_0 \in \mathcal{P}_4(\mathbb{R}^3)$.

D'après le lemme 3.13 :

$$M := \sup_{t \geq 0} \int_{\mathbb{R}^3} |v|^4 \, df(t)(v) < +\infty.$$

Cette condition implique en fait que la famille de probabilités $(f(t))_{t \in \mathbb{R}^+}$ est tendue. En effet, si $t \geq 0$ et si $R > 0$:

$$\begin{aligned}
f(t)(B(0,R)^c) &= \int_{|v| \geq R} df(t) \\
&\leq \int_{|v| \geq R} \frac{|v|^4}{R^4} \, df(t)(v) \\
&\leq \frac{1}{R^4} \int_{|v| \geq R} |v|^4 \, df(t)(v) \\
&\leq \frac{M}{R^4}.
\end{aligned}$$

D'où la tension de $(f(t))_{t\in\mathbb{R}^+}$. L'inégalité de Cauchy-Schwarz implique que :

$$\sup_{t\geq 0}\int_{|v|\geq R}|v|^2\,df(t)(v) \underset{R\to+\infty}{\to} 0.$$

Le théorème 1.8 de Prokhorov et le théorème 2.10 sur la convergence W_2 assure qu'il existe une suite $(t_n)_n$ strictement croissante et une loi $\mu_0 \in \mathcal{P}(\mathbb{R}^3)$ telles que :

$$W_2(f(t_n),\mu_0)\to 0.$$

On remarque que μ_0 a les mêmes deux premiers moments que f_0 et donc que G. Montrons que $\mu_0 = G$. Puisque G est un équilibre de (B), le théorème 3.9 permet d'affirmer que :

$$W_2(f(t_{n+1}),G)\leq W_2(f(t_n+1),G)\leq W_2(f(t_n),G)\quad (*).$$

Par encadrement, on déduit que $W_2(f(t_n+1),G) \underset{n\to+\infty}{\to} W_2(\mu_0,G)$. On note $\mu(t)$ la solution de (B) de donnée initiale μ_0. Alors :

$$W_2(f(t_n+1),\mu(1))\leq W_2(f(t_n),\mu_0) \quad \underset{n\to+\infty}{\to} 0.$$

Par inégalité triangulaire, on conclut que :

$$W_2(f(t_n+1),G) \underset{n\to+\infty}{\to} W_2(\mu(1),G).$$

Passant à la limite dans $(*)$, on a :

$$W_2(\mu(1),G) = W_2(\mu_0,G)$$

puis

$$\forall t\in[0,1],\quad W_2(\mu(t),G)=W_2(\mu_0,G)$$

d'après le théorème 3.9. La distance entre les deux solutions $\mu(t)$ et G est donc constante sur $[0,1]$. En étudiant le cas d'égalité dans la preuve du théorème 3.9, on doit avoir :

$$\forall t\in[0,1],\quad W_2(Q_+(\mu(t)),Q_+(G)) = W_2(\mu(t),G).$$

Or $\mu(t)$ et G ont les mêmes deux premiers moments d'après la deuxième partie du théorème 3.7, on peut appliquer le lemme 3.10. On a donc $\mu_0 = G$. Finalement on a montré : $W_2(f(t_n),G) \underset{n\to+\infty}{\to} 0$ et donc $W_2(f(t),G) \underset{t\to+\infty}{\to} 0$ car il s'agit d'une fonction décroissante d'après le théorème 3.9.

Étape 2 : On se ramène au cas précédent par troncature.

Soit $f_0 \in \mathcal{P}_2(\mathbb{R}^3)$, on considère une suite $(f_0^n)_n$ de $\mathcal{P}_4(\mathbb{R}^3)$ ayant les mêmes deux premiers moments que f_0 et qui converge étroitement vers f_0. On peut construire une telle suite par troncature pour que le moment d'ordre 4 soit fini, puis par translation et dilatation pour préserver les deux premiers moments. Le théorème 2.10 qui caractérise la convergence en distance de Wasserstein assure que :

$$W_2(f_0^n,f_0)\to 0.$$

De plus, on a pour tout $t \in \mathbb{R}^+$ et pour tout $n \in \mathbb{N}$:

$$\begin{aligned} W_2(f(t), G) &\leq W_2(f(t), f^n(t)) + W_2(f^n(t), G) \qquad \text{(par inégalité triangulaire)} \\ &\leq W_2(f_0, f_0^n) + W_2(f^n(t), G) \quad \text{(d'après le théorème 3.9)} . \end{aligned}$$

On fixe $\epsilon > 0$, il existe $N \in \mathbb{N}$ tel que $W_2(f_0, f_0^N) \leq \epsilon$. Comme $f_0^N \in \mathcal{P}_4(\mathbb{R}^3)$, il existe $T > 0$ tel que :

$$\forall t \geq T, \quad W_2(f^N(t), G) \leq \epsilon,$$

d'après l'Étape 1. Ainsi, on a :

$$\forall t \geq T, \quad W_2(f(t), G) \leq 2\epsilon.$$

□

Remarque 3.15. *Ce théorème montre également que les seuls équilibres de l'équation de Boltzmann sont les maxweliennes. De plus, comme les maxwelliennes ont une matrice de covariance diagonale, on déduit qu'asymptotiquement, les vitesses selon les trois directions sont indépendantes.*

4 Annexe : définition d'une probabilité par dualité

Le théorème fondamental qui permet de définir une probabilité, ou plus généralement une mesure de Borel à partir d'une forme linéaire est le théorème de Riesz.

Théorème 4.1 (Théorème de Riesz pour les mesures)**.**
Soit (X,d) un espace métrique localement compact et séparable et soit ϕ une forme linéaire positive sur $C_c(X,\mathbb{R})$. C'est-à-dire que :

$$\forall f \in C_c(X,\mathbb{R}^+), \quad \phi(f) := \langle \phi, f \rangle \geq 0.$$

Alors il existe une unique mesure sur la tribu borélienne $B(X)$ telle que :

$$\forall f \in C_c(X,\mathbb{R}), \quad \langle \phi, f \rangle = \int_X f \, d\mu.$$

De plus μ est une mesure de Borel, i.e. une mesure positive sur $\mathcal{B}(X)$, et finie sur les compacts. On a également :

$$\forall \Omega \in \mathcal{O}(X), \quad \mu(\Omega) = \sup \left\{ \langle \phi, f \rangle, \, f \in C_c(X,[0,1]), f \leq \mathbb{1}_\Omega \right\}.$$

Ce théorème permet notamment de définir une probabilité par dualité en définissant une forme linéaire positive sur $C_c(X,\mathbb{R})$.

Définition 4.2.
Soient (X,d) un espace polonais localement compact et (T,λ) un espace de probabilité. On se donne également une application mesurable :

$$\left\{ \begin{array}{ccc} T & \to & \mathcal{P}(X) \\ t & \mapsto & \mu(t) \end{array} \right. .$$

On définit par dualité la probabilité M par :

$$\forall f \in C_c(X,\mathbb{R}), \quad \langle M, f \rangle = \int_T \left(\int_X f(x) \, d\mu(t)(x) \right) d\lambda(t),$$

on note alors :

$$M := \int_T \mu(t) \, d\lambda(t).$$

Le fait qu'on définit bien une mesure positive est assuré par le théorème 4.1 de Riesz. Pour voir qu'il s'agit bien d'une probabilité, il suffit de prendre une suite croissante $(f_n)_n$ de $C_c(X,[0,1])$ qui converge simplement vers la fonction constante égale à 1. Alors :

$$\forall t \in T, \quad \int_X f_n \, d\mu(t) \underset{n \to +\infty}{\nearrow} 1.$$

Le théorème de convergence monotone assure que :

$$\langle M, f_n \rangle \to 1,$$

ce qui montre que :

$$M(X) = \sup \left\{ \langle M, f \rangle, \, f \in C_c(X,[0,1]) \right\} = 1.$$

Références

[1] C. Villani, *Topics on optimal transportation*, American Mathematical Society, 2003.

[2] C. Villani, Optimal transport. *Old and new.* Grundlehren der Mathematischen Wissenschaften, volume 338. Springer-Verlag, Berlin, 2009.

[3] R.M. Dudley, *Real analysis and probability*, Cambridge University Press, 2002.

[4] G. Royer, *An initiation to logarithmic Sobolev inequalities*, American Mathematical Society, 2007.

[5] P. A. Markowich et C. Villani., *On the trend to equilibrium for the Fokker-Planck equation : an interplay between physics and functional analysis*, VI Workshop on Partial Differential Equations, Part II (Rio de Janeiro, 1999). Mat. Contemp. 19 (2000), 1–29.

[6] C. Villani, *A review of mathematical topics in collisional kinetic theory*, 2002, `http://cedricvillani.org/wp-content/uploads/2012/08/preprint.pdf`.