Théorème centraux limites pour le coût de transport empirique.

Robinson Le Bihan

Encadré par Hélène Guérin et Dinh-Toan Nguyen

2 Mai - 30 Juin 2023

Table des matières

1 Introduction

Le transport optimal de mesures est un domaine des mathématiques dont l'origine remonte au XVIIIème siècle avec le mathématicien Français Gaspard Monge qui a formulé en 1781 le problème, aujourd'hui connu comme "problème de Monge" qui consiste à déterminer comment déplacer une certaine quantité de ressource d'une configuration initiale - formalisé mathématiquement par une mesure - à une configuration finale - une autre mesure - de manière optimale, en minimisant le coût total de transport. Ce problème peut prendre de nombreuses formes, on peut s'imaginer vouloir déplacer un tas de sable, réordonner des livres dans une bibliothèque, organiser le transport de fer entre les mines et les usines et autres situations. Cependant ce problème est compliqué à la résolution et est plutôt mal posé dans le sens où il n'existe pas nécessairement de processus de transport optimal. Il a alors fallu attendre le XXème siècle pour qu'en 1942 les travaux du mathématicien russe Leonid Kantorovich apportent une formulation plus générale du problème, appelé "problème de Monge-Kantorovich" formalisée différemment mathématiquement mais bien plus confortable car d'une part il existe des solutions, le problème est symétrique et d'autre part de nombreux nouveaux outils s'ajoutent grâce à cette approche. Kantorovich a ensuite reçu en 1975 le prix Nobel d'économie pour ses travaux sur la théorie du transport optimal et ses applications en économie. Le transport optimal de mesures a ensuite été étudié sous de nombreux aspects mathématiques, aux applications nombreuses (économie, physique par exemple).

Le transport optimal de mesures permet de comparer des mesures de probabilités, en particulier la distance de Wassertein est une distance sur l'espace des mesures de probabilités (sur $\mathbb{R}^d$ par exemple) définie à partir du coût de transport optimal. On peut ensuite imaginer une démarche approchant des problèmes de statistiques et étudier la distance de Wasserstein entre une mesure de probabilité empirique qui représente une expérience statistique et une mesure de probabilité fixe. La question posée peut par exemple être "combien coûte le transport d'une mesure de probabilité empirique vers la loi sous-jacente ?". On se doute, et c'est vérifié (nous le verrons), que ce coût de transport - par la distance de Wasserstein - converge vers 0. On peut aussi comparer deux mesures empiriques différents, c'est-à-dire étudier un coût de correspondance optimal.

Aux alentours de 2017 Eustasio Del Barrio (mathématicien espagnol) et Jean-Michel Loubes (mathématicien français) étudiaient la question du coût de transport optimal entre une mesure empirique et une mesure cible et ont approfondit la convergence vers 0 en apportant une vitesse de convergence et le comportement probabiliste des fluctuations autours de la moyenne, c'est-à-dire qu'ils sont parvenus à ériger un théorème limite central.

2 Premières notions de transport optimal de mesures

Les résultats de cette section proviennent en partie des chapitres 1, 2 et 7 de [GSZ18].

2.1 Problème de Monge

La question posée par Gaspard Monge en 1781 est la suivante : comment déplacer de manière optimale un tas de sable dans un trou ? La première hypothèse effectuée est que le tas de sable remplit le trou, c'est à dire qu'ils occupent le même volume. Ensuite, il faut comprendre ce qu'on entend par une manière optimale de déplacer le tas de sable : le faire de la manière la moins coûteuse. Par exemple si l'on considère déplacer un grain x vers un grain cible y dans le trou coûte leur distance $||x-y||$, optimiser le transport du tas de sable reviendrait donc à minimiser le coût total de déplacement de tous les grains de sable, par rapport à tous les déplacements possibles.

Concrètement, on va modéliser le tas de sable à déplacer par une mesure μ et le trou par une "mesure cible" ν, définies sur $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Comme expliqué on exige que les trous de même volume, on va donc exiger également que ce sont des mesures de probabilités. Transporter μ sur ν veut dire trouver une application mesurable $T : \mathbb{R}^d \to \mathbb{R}^d$ qui transporte μ sur ν, dans le sens où chaque sous-partie du trou ν, donc tout $B \in \mathscr{B}(\mathbb{R}^d)$ (quitte à prendre les négligables) a le même volume que la sous-partie du tas de sable constituée des grains de sables qui sont envoyés dans B. On demande donc

$$\forall B \in \mathscr{B}(\mathbb{R}^d), \mu\left(\left\{x \in \mathbb{R}^d ; T(x) \in B\right\}\right) = \nu(B)$$

c'est-à-dire que la mesure image de μ par T est égale à la mesure cible ν, où la mesure image de μ par T est définie par

$$T_{\#\mu} : B \in \mathscr{B}(\mathbb{R}^d) \to \mu\left(T^{-1}(B)\right).$$

A propos de l'optimalité, en restant dans le cas général, on définit une fonction de coût par une fonction mesurable $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+ \cup \{+\infty\}$, (par exemple distance euclidienne) où donc $c(x,y)$ représente ce que coûte le déplacement du grain de sable x vers la position y. Le coût de transport via l'application T est donc la quantité

$$\int_{\mathbb{R}^d} c(x, T(x)) d\mu(x),$$

que l'on cherche à minimiser. Le problème de Monge est donc le suivant : étant donné deux mesures de probabilités μ et ν sur les boréliens de $\mathbb{R}^d$, on cherche, si elle existe, une application T qui transporte μ sur ν, telle que

$$\int_{\mathbb{R}^d} c(x, T(x)) d\mu(x) = \inf\left\{\int_{\mathbb{R}^d} c(x, T(x)) d\mu(x);\ T : \mathbb{R}^d \to \mathbb{R}^d \text{ est mesurable et vérifie } T_{\#\mu} = \nu\right\}.$$

Remarque 2.1. Le cadre proposé permet l'utilisation du théorème de transfert. Ainsi si μ et ν sont ses mesures de probabilités sur $\mathbb{R}^d$ et $T_{\#\mu} = \nu$ alors

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

2.2 Problème de Monge-Kantorovich

Cependant le problème de Monge est mal posé dans le sens où de nombreux exemples prouvent qu'il n'existe pas forcément d'application de transport, qu'elle n'est pas toujours unique quand elle existe ou alors qu'une ou plusieurs peuvent exister sans qu'aucune ne soit optimale. Léonard Kantorovich a alors apporté une version généralisée - on verra en quoi - du problème de Monge, plus satisfaisante. Appuyons son idée par un exemple : sur $\mathbb{R}$, on ne peut pas transporter la masse δ_0 sur $\frac{\delta_{-1}+\delta_1}{2}$. En effet si T réalise le transport entre ces deux mesures, c'est-à-dire que $T_{\#\delta_0} = \frac{\delta_{-1}+\delta_1}{2}$, donc $T_{\#\delta_0}(\{1\}) = 1/2$, or $T_{\#\delta_0}(\{1\}) = \delta_0(T^{-1}(\{1\}) \in \{0,1\}$, donc T ne peut pas exister. Pour transporter δ_0 sur $\frac{\delta_{-1}+\delta_1}{2}$ il faudrait s'autoriser de framenter un grain de sable en deux demis grains. L'idée de Kantorovich est en fait, en plus de disperser le tas de sable, de disperser chaque grain de sable. Concrètement, à chaque point/grain de sable $x \in \mathbb{R}^d$ on associe une mesure de probabilité p_x/manière de l'éparpiller. On définit en particulier la notion de noyau.

Définition 2.2 (Noyau de transition)**.** Un noyau de probabilité sur $\mathbb{R}^d$ est une application p qui à chaque x associe une mesure de probabilité p_x sur $\mathbb{R}^d$. On exige de plus que pour toute fonction test $h : \mathbb{R}^d \to \mathbb{R}$ mesurable bornée, la fonction $x \mapsto \displaystyle\int_{\mathbb{R}^d} h(y)dp_x(y)$ est mesurable.

Le processus de transport correspond donc à tirer un $x \in \mathbb{R}^d$ selon la loi μ puis conditionnellement à ce tirage tirer un $y \in \mathbb{R}^d$ selon la loi p_x, c'est-à-dire que pour transporter μ vers ν au sens de Kantorovich, il faudra exiger

$$\forall B \in \mathscr{B}(\mathbb{R}^d), \nu(B) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathbf{1}_B(y)dp_x(y) \right) d\mu(x) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{1}_B(y)dp_x(y)d\mu(x).$$

et le coût de transport de μ à ν devient alors la quantité

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)d\mu(x)dp_x(y), \tag{1}$$

toutes ces quantités ayant un sens si $x \mapsto p_x$ est un noyau. Introduisons la notion de couplage, dont l'intérêt sera révélé instantanément :

Définition 2.3. Un couplage π de μ et de ν est une mesure de probabilités sur $\mathbb{R}^d \times \mathbb{R}^d$ dont les lois marginales sont pour la première μ et pour la deuxième ν, c'est-à-dire que pour toute fonction $\varphi : \mathbb{R}^d \to \mathbb{R}$ (resp. ψ) μ-mesurable (resp. ν-mesurable),

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x)d\pi(x,y) = \int_{\mathbb{R}^d} \varphi(x)d\mu(x) \ \left(\text{resp. } \iint_{\mathbb{R}^d \times \mathbb{R}^d} \psi(y)d\pi(x,y) = \int_{\mathbb{R}^d} \psi(y)d\nu(y) \right).$$

Nous noterons $\Pi(\mu,\nu)$ l'ensemble des couplages de μ et ν.

Nous admettrons le théorème suivant, qui explique l'utilité des couplages et simplifie l'énoncé du problème de Kantorovich. Il correspond au théorème de Jirina d'existence de lois conditionnelles (voir [Jiř54])

Théorème 2.4. *Soit $\pi \in \Pi(\mu,\nu)$. Il existe un noyau de transition p tel que pour toute fonction π-mesurable bornée $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$,*

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} h(x,y)d\pi(x,y) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} h(x,y)dp_x(y) \right) d\mu(x),$$

c'est-à-dire que $\pi(dxdy) = \mu(dx)p_x(dy)$. p est μ-presque sûrement unique, dans le sens où si p' est une autre noyau de transition qui vérifie ce théorème, $p = p'$ μ-presque sûrement. Réciproquement pour tout noyau de transition p, $\mu(dx)p_x(dy)$ est un couplage.

Ainsi, transporter μ sur ν au sens de Kantorovich revient à trouver un couplage π Il convient alors de définir la notation utilisée pour le coût de transport d'un couplage :

Définition 2.5 (Coûts de transport pour le problème de Kantorovich)**.** Soit $\pi \in \Pi(\mu, \nu)$, le coût de transport associé à π (relativement à c) est la quantité

$$I_c(\pi) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi(x, y). \tag{2}$$

Le coût de transport optimal (sous réserve d'existence) de μ à ν est la quantité

$$T_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} I_c(\pi). \tag{3}$$

Un couplage π^* réalisant le coût de transport optimal sera appelé un plan de transport optimal.

Le problème de Monge-Kantorovich est donc le suivant : étant donné μ et ν, minimiser le coût de transport et trouver un plan de transport optimal.

Remarque 2.6. Le problème de Monge-Kantorovich généralise le problème de Monge. En effet s'il existe une application de transport T de μ à ν alors $\pi_T := \mu(dx)\delta_{T(x)}(dy)$ est un couplage, donc de noyau $\delta_{T(\cdot)}$. Ainsi,

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi_T(x, y) = \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x).$$

Egalement, si π^* est un plan de transport optimal sous la forme π_T, alors l'application T correspondante est solution du problème de Monge. Un tel couplage est appelé un couplage déterministe

Au-delà de généraliser le problème de Monge on verra que le problème de Monge-Kantorovich est plus confortable dans le sens où il fournit l'existence de plans de transports optimaux dès que la fonction de coût est un peu régulière. Il y aura unicité et déterminisme dans certains cas qui seront décris par le théorème 4.2 de Brenier.

2.3 Résolution du problème de Monge-Kantorovich - Existence de plans de transport optimaux

2.3.1 Topologie de la convergence étroite sur l'ensemble des mesures de probabilités

Munissons l'ensemble des mesures de probabilités sur $\mathbb{R}^d$ d'une certaine topologie.

Définition 2.7 (Topologie de la convergence étroite)**.** Un ensemble U de mesures de probabilités sur $\mathbb{R}^d$ est un ouvert pour la topologie de la convergence étroite si pour tout mesure de probabilités μ_0 sur $\mathbb{R}^d$ il existe $p \in \mathbb{N}^*$, $\varphi_1, \ldots, \varphi_p$ continues et bornées sur $\mathbb{R}^d$ et $r_1, \ldots, r_p > 0$ tels que

$$\bigcap_{i=1}^{p} \left\{ \mu \text{ probabilité sur } \mathbb{R}^d ; \left| \int_{\mathbb{R}^d} \varphi_i d\mu - \int_{\mathbb{R}^d} \varphi_i d\mu_0 \right| < r_i \right\} \subset U.$$

Remarque 2.8. La topologie de la convergence étroite est la topologie la moins fine qui rend continues toutes les applications $\mu \mapsto \int_{\mathbb{R}^d} \varphi d\mu$ quand $\varphi : \mathbb{R}^d \to \mathbb{R}$ est continue et bornée. Une suite de mesures de probabilités $(\mu_n)_n$ converge étroitement vers une autre mesure de probabilités μ si

$$\text{pour toute } \varphi : \mathbb{R}^d \to \mathbb{R} \text{ continue et bornée, } \int_{\mathbb{R}^d} \varphi d\mu_n \underset{n \to +\infty}{\longrightarrow} \int_{\mathbb{R}^d} \varphi d\mu.$$

et on notera

$$\mu_n \xrightarrow[n\to+\infty]{w} \mu. \tag{4}$$

Une suite de variables aléatoires $(X_n)_n$ converge en loi si et seulement si la suite des lois des X_n converge étroitement.

Théorème 2.9. *La topologie de la convergence étroite est métrisable.*

Un exemple de métrique pour la topologie de la convergence étroite est celle de Prokhorov-Lévy, voir [Bil99]. Ceci permet une approche plus confortable par les suites de la notion de compacité qui sera utile. En particulier, la question de la compacité pour cette topologie est abordée dans le théorème suivant, admis :

Théorème 2.10. *Soit A un ensemble de mesures de probabilités sur $\mathbb{R}^d$. A est d'adhérence compacte si et seulement si il est uniformément tendu, c'est-à-dire que pour tout ε il existe un compact K_ε de $\mathbb{R}^d$ tel que*

$$\sup_{\mu\in A} \mu\left(\mathbb{R}^d \setminus K_\varepsilon\right) \leq \varepsilon.$$

2.3.2 Semi-continuité inférieure

Pour prouver l'existence de plan de transport optimal, on va avoir besoin de la notion de semi-continuité inférieure. En effet, on va appliquer un argumentaire type "fonction continue sur un compact et minorée atteint sa borne inférieure". On réussira en partie 2.3.3 à prouver que quand c est continue $I_c : \pi \mapsto I_c(\pi)$ est une fonction semi-continue inférieurement sur l'ensemble des mesures de probabilités sur $\mathbb{R}^d \times \mathbb{R}^d$. On prouvera dans cette partie qu'une fonction semi-continue inférieurement sur un compact atteint sa borne inférieure. Le théorème 2.10 permettra alors de montrer que $\Pi(\mu, \nu)$ est un compact pour la topologie de la convergence étroite, ainsi I_c y atteindra sa borne inférieure, il existera donc un plan de transport optimal. Soit (E, d) un espace métrique.

Définition 2.11 (Semi-continuité inférieure)**.** Soit $f :\to \mathbb{R}\cup\{+\infty\}$ et $x \in E$. f est semi-continue inférieurement en x si pour tout suite $(x_n)_n \subset E$ convergent vers x,

$$\liminf_{n\to+\infty} f(x_n) \geq f(x).$$

Remarque 2.12. A partir de maintenant on n'écrira plus "semi-continue inférieurement" mais s.c.i comme abréviation.

Proposition 2.13. *Soit $f : E \to \mathbb{R}\cup\{+\infty\}$.*

1. *f est s.c.i sur E si et seulement si pour tout $r \in \mathbb{R}$ l'ensemble $f^{-1}(]-\infty, r])$ est fermé.*
2. *si f est s.c.i sur un compact $K \subset E$ elle y est minorée et y atteint sa borne inférieure.*

Preuve. 1. On suppose que f est s.c.i. Soit $r \in \mathbb{R}$, montrons que $f^{-1}(]-\infty, r])$ est fermé. Soit donc $(x_n)_n \subset f^{-1}(]-\infty, r])$ qui converge vers $x \in E$, alors pour tout $n \in \mathbb{N}$, $f(x_n) \leq r$. f étant s.c.i en x, $f(x) \leq \liminf_{n\to+\infty} f(x_n) \leq r$ donc $x \in f^{-1}(]-\infty, r])$ donc $f^{-1}(]-\infty, r])$ est fermé.
On suppose réciproquement que $f^{-1}(]-\infty, r])$ est fermé. Soit $(x_n)_n \subset E$ convergeant vers $x \in \mathbb{R}^d$ et on suppose d'abord que $f(x) < +\infty$. Soit $\varepsilon > 0$, donc $f^{-1}(]f(x)-\varepsilon, +\infty[)$ est un ouvert qui contient x. Alors à partir d'un certain rang $N \in \mathbb{N}$, pour $n \geq N$, $x_n \in f^{-1}(]f(x)-\varepsilon, +\infty[)$ c'est-à-dire que $f(x_n) > f(x)-\varepsilon$ donc en passant à la limite inférieure $\liminf f(x_n) > f(x)-\varepsilon$. On conclut en faisant

tendre ε vers 0. Si $f(x)=+\infty$ on conclut avec les ouverts de la forme $f^{-1}(]M,+\infty[)$.

2. On suppose par l'absurde que f n'est pas minorée sur K. : il existe une suite $(x_n)_n \subset K$ telle que $f(x_n) \to -\infty$. Par compacité on peut extraire $(x_{\varphi(n)})_n$ telle que $x_{\varphi(n)} \to x \in K$. f étant s.c.i, $-\infty = \liminf f(x_{\varphi(n)}) \geq f(x)$, absurde donc f est minorée et $m := \inf_{x\in K} f(x) \in \mathbb{R}$ (on suppose que f n'est pas identiquement $+\infty$ par ailleurs). Soit maintenant $(x_n) \subset K$ telle que $f(x_n) \to m$, par compacité on extrait $(x_{\varphi(n)})_n$ qui converge vers $x \in K$. f étant s.c.i, $m = \liminf f(x_n) \geq f(x) \geq m$ donc $f(x) = m$. □

2.3.3 Existence de plan de transport optimal

Proposition 2.14. *Soient μ et ν deux mesures de probabilités sur $\mathbb{R}^d$. Alors $\Pi(\mu,\nu)$ est un compact dans l'ensemble des mesures de probabilités sur $\mathbb{R}^d \times \mathbb{R}^d$ pour la topologie de la convergence étroite.*

Preuve. Montrons d'abord $\Pi(\mu,\nu)$ est un ensemble fermé. Soit $(\pi_n)_n$ une suite de couplages qui converge étroitement vers une mesure π sur $\mathbb{R}^d \times \mathbb{R}^d$. π est bien une mesure de probabilités en prenant l'application $x \in \mathbb{R}^d \mapsto 1$, c'est-à-dire

$$1 = \int_{\mathbb{R}^d} 1 d\pi_n(x,y) \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} d\pi(x,y).$$

Montrons ensuite que π a pour marginales μ et ν. Soit $\varphi : \mathbb{R}^d \to \mathbb{R}$ continue et bornée, étant donné que π_n a μ pour première marginale,

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(x) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x) d\pi_n(x,y) \underset{n\to+\infty}{\longrightarrow} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x) d\pi(x,y)$$

donc π a pour première marginale μ. De même π a pour deuxième marginale ν. On a bien montré que $\Pi(\mu,\nu)$ est fermé pour la topologie de la convergence étroite. il ne reste plus qu'à montrer que c'est un ensemble relativement compact. D'après le théorème 2.10 il s'agit de montrer que $\Pi(\mu,\nu)$ est uniformément tendu. Soit $\varepsilon > 0$ fixé. Les ensembles $\{\mu\}$ et $\{\nu\}$ sont compacts donc d'après le théorème 2.10 il existe $K_{\varepsilon,1}$ et $K_{\varepsilon,2}$ tels que $\mu(\mathbb{R}^d \setminus K_{\varepsilon,1}) \leq \varepsilon$ et $\nu(\mathbb{R}^d \setminus K_{\varepsilon,2}) \leq \varepsilon$. Posons alors $K_\varepsilon = K_{\varepsilon,1} \times K_{\varepsilon,2}$ qui est un compact de $\mathbb{R}^d \times \mathbb{R}^d$ comme produit de compacts. Pour $\pi \in \Pi(\mu,\nu)$,

$$\begin{aligned}\pi(\mathbb{R}^d \times \mathbb{R}^d \setminus K_\varepsilon) &= \pi\left(\left((\mathbb{R}^d \setminus K_{\varepsilon,1}) \times \mathbb{R}^d\right) \cup \left(\mathbb{R}^d \times (\mathbb{R}^d \setminus K_{\varepsilon,2})\right)\right) \\ &\leq \pi\left((\mathbb{R}^d \setminus K_{\varepsilon,1}) \times \mathbb{R}^d\right) + \pi\left(\mathbb{R}^d \times (\mathbb{R}^d \setminus K_{\varepsilon,2})\right) = \mu\left(\mathbb{R}^d \setminus K_{\varepsilon,1}\right) + \nu\left(\mathbb{R}^d \setminus K_{\varepsilon,2}\right) \leq 2\varepsilon\end{aligned}$$

on déduit bien que $\Pi(\mu,\nu)$ est uniformément tendu, donc relativement compact et compact car fermé.
□

On va pouvoir prouver le théorème d'existence de plan de transports optimaux.

Théorème 2.15. *Soit $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+ \cup \{+\infty\}$ une fonction de coût continue. Pour toutes μ,ν mesures de probabilités sur $\mathbb{R}^d$ il existe un couplage $\pi^* \in \Pi(\mu,\nu)$ optimal, c'est-à-dire tel que (se référer à (2) et (3) pour les notations)*

$$I_c(\pi^*) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y) d\pi^*(x,y) = T_c(\mu,\nu).$$

Preuve. Montrons que la fonction $I_c : \pi \mapsto I_c(\pi)$ est s.c.i sur l'ensemble des mesures de probabilités sur $\mathbb{R}^d \times \mathbb{R}^d$ pour la topologie de la convergence faible, qui est un espace métrique (proposition 2.9). Pour $n \in \mathbb{N}$ posons $c_n = \min\{c, n\} = c$ quand $c < n$ et n sinon, ainsi $(c_n)_n$ est une suite croissante de fonctions continues et bornées. Soit $\pi \in \Pi(\mu, \nu)$. Par convergence monotone

$$I_{c_n}(\pi) = \int_{\mathbb{R}^d \times \mathbb{R}^d} c_n(x, y) d\pi(x, y) \underset{n \to +\infty}{\longrightarrow} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi(x, y) = I_c(\pi).$$

Invoquons le point 1. de la proposition 2.13 pour montrer que I_c est s.c.i. Soit $r \in \mathbb{R}$, alors $I_c(\pi) \leq r$ si et seulement si pour tout $n \in \mathbb{N}$, $I_{c_n}(\pi) \leq r$ donc $I_c(]-\infty, r]) = \bigcap_{n \in \mathbb{N}} I_{c_n}(]-\infty, r])$ est un fermé comme intersection de fermés. Ainsi par la proposition 2.13, I_c est s.c.i, en particulier elle l'est sur $\Pi(\mu, \nu)$ qui est un compact et est minorée donc elle y atteint sa borne inférieure. $\square$

Ainsi nous avons prouvé qu'il existe toujours un plan de transport optimal dès lors que la fonction de coût est continue. Ainsi le problème de Monge-Kantorovich généralise le problème de Monge et est mieux posé en ce qu'il existe - avec la continuité - toujours une solution. L'unicité n'est pas forcément acquise mais, notre étude se recentrant ensuite au cas de la fonction de coût $(x, y) \mapsto ||x - y||^2$ associée à la distance quadratique, nous verrons qu'il est unique et même déterministe, c'est-à-dire qu'il existe une unique solution au problème de Monge.

Remarque 2.16. Tout ce qui a été énoncé pour le moment est adaptable sur un espace métrique polonais (séparable et complet) quelconque. Egalement, l'hypothèse de continuité sur c peut être affaible, en effet il suffit que la fonction de coût soit s.c.i pour que I_c soit s.c.i et ainsi qu'il existe un plan de transport optimal. Ce n'est qu'évoqué car dans le cadre particulier qui nous intéressera par la suite (coût quadratique) ce n'est pas utile.

2.4 Distance de Wasserstein

Le coût de transport entre deux mesures de probabilités peut s'interpréter de manière heuristique comme une manière d'appréhender la "proximité" entre ces deux mesures : plus il est faible, plus ces deux mesures seraient proches. Approfondir ce point de vue revient à se demander si le coût de transport est une distance.

Définition 2.17. On note pour $p \geq 1$, $\mathscr{P}_p(\mathbb{R}^d)$ l'ensemble des mesures de probabilités sur les boréliens de $\mathbb{R}^d$ qui admettent un moment d'ordre p :

$$\mathscr{P}_p(\mathbb{R}^d) = \left\{ \mu \text{ mesure de probabilités sur } \mathbb{R}^d ; \int_{\mathbb{R}^d} ||x||^p d\mu(x) \right\}.$$

et pour $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$ on note $T_p(\mu, \nu)$ le coût de transport entre μ et ν pour la fonction de coût $(x, y) \mapsto ||x - y||^p$ ainsi que

$$W_p(\mu, \nu) := T_p(\mu, \nu)^{\frac{1}{p}} = \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d} ||x - y||^p d\pi(x, y) \right)^{\frac{1}{p}}.$$

Proposition 2.18. *Définie ainsi, W_p est une distance sur $\mathscr{P}_p(\mathbb{R}^d)$.*

Preuve. Soient $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$. Montrons déjà que $W_p(\mu,\nu) < +\infty$. En effet si $\pi \in \Pi(\mu,\nu)$,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} ||x-y||^p d\pi(x,y) \leq \int_{\mathbb{R}^d\times\mathbb{R}^d} (||x||+||y||)^p d\pi(x,y) \leq \int_{\mathbb{R}^d\times\mathbb{R}^d} (2\max\{||x||,||y||\})^p d\pi(x,y)$$
$$= 2^p \int_{\mathbb{R}^d\times\mathbb{R}^d} (\max\{||x||,||y||\})^p d\pi(x,y) \leq 2^p \left(\int_{\mathbb{R}^d} ||x||^p d\mu(x) + \int_{\mathbb{R}^d} ||y||^p d\nu(y) \right) < +\infty.$$

La **positivité** est immédiate, montrons la **symétrie** :

$$W_p(\mu,\nu) = \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d} ||x-y||^p d\pi(x,y) \right)^{\frac{1}{p}} = \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d} ||y-x||^p d\pi(x,y) \right)^{\frac{1}{p}} = W_p(\nu,\mu).$$

Montrons la **séparation** : supposons maintenant que $W_p(\mu,\nu) = 0$ alors, la fonction de coût étant continue, par le théorème 2.15 il existe un plan de transport optimal π^*. Ainsi

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} ||x-y||^p d\pi^*(x,y) = 0$$

donc par positivité $x = y$ pour π^*-presque tout (x,y). Donc pour $B \in \mathscr{B}(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \mathbf{1}_B(x) d\mu(x) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{1}_B(x) d\pi^*(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{1}_B(y) d\pi^*(x,y) = \int_{\mathbb{R}^d} \mathbf{1}_B(y) d\nu(y).$$

donc $\mu = \nu$. Montrons que $W_p(\mu,\mu) = 0$, introduisons $T : x \in \mathbb{R}^d \mapsto (x,x)$ et $\pi = T_{\#\mu}$ alors $\pi \in \Pi(\mu,\mu)$ et par le théorème de transfert,

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} ||x-y||^p d\pi(x,y) = \int_{\mathbb{R}^d} ((x,y) \mapsto ||x-y||^p) \circ T(x) d\mu(x) = \int_{\mathbb{R}^d} ||x-x||^p d\mu(x) = 0$$

donc, la quantité sur laquelle la borne inférieure est prise étant positive, alors $W_p(\mu,\mu) = 0$. Reste à montrer **l'inégalité triangulaire** mais énonçons d'abord un lemme admis, démontré dans [Vil03].

Lemme 2.19 (Recollement)**.** *Soient μ_1, μ_2 et μ_3 des mesures de probabilités sur $\mathbb{R}^d$, $\pi_{1,2} \in \Pi(\mu_1,\mu_2)$, $\pi_{2,3} \in \Pi(\mu_2,\mu_3)$ des plans de transport optimaux, alors il existe une mesure de probabilités π sur $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ qui possède $\pi_{1,2}$ pour marginale vis-à-vis des deux premières variables et $\pi_{2,3}$ vis-à-vis des deux dernières variables.*

Soient désormais $\mu_1, \mu_2, \mu_3 \in \mathscr{P}_p(\mathbb{R}^d)$, alors si π est la mesure évoquée dans le lemme de recollement 2.19,

$$\left(\int_{(\mathbb{R}^d)^3} ||x_1-x_3||^p d\pi(x_1,x_2,x_3) \right)^{\frac{1}{p}} \leq \left(\int_{(\mathbb{R}^d)^3} (||x_1-x_2||+||x_2-x_3||)^p d\pi(x_1,x_2,x_3) \right)^{\frac{1}{p}}$$
$$= \Big|\Big| (x_1,x_2,x_3) \mapsto ||x_1-x_2||+||x_2-x_3|| \Big|\Big|_{L^p(\pi)} \leq \Big|\Big| (x_1,x_2,x_3) \mapsto ||x_1-x_2|| \Big|\Big|_{L^p(\pi)}$$
$$+ \Big|\Big| (x_1,x_2,x_3) \mapsto ||x_2-x_3|| \Big|\Big|_{L^p(\pi)}$$

La dernière inégalité provenant de l'inégalité triangulaire sur $L^p(\pi)$. Par le lemme 2.19,

$$\Big|\Big| (x_1,x_2,x_3) \mapsto ||x_1-x_2|| \Big|\Big|_{L^p(\pi)} = \left(\int_{(\mathbb{R}^d)^3} ||x_1-x_2||^p d\pi(x_1,x_2,x_3) \right)^{\frac{1}{p}}$$
$$= \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} ||x_1-x_2||^p d\pi_{1,2}(x_1,x_2) \right)^{\frac{1}{p}} = W_p(\mu_1,\mu_2). \quad (5)$$

et de même $\left|\left|(x_1, x_2, x_3) \mapsto ||x_2 - x_3||\right|\right|_{L^p(\pi)} = W_p(\mu_2, \mu_3)$ et par définition

$$W_p(\mu_1, \mu_3) \leq \left(\int_{(\mathbb{R}^d)^3} ||x_1 - x_3||^p d\pi(x_1, x_2, x_3) \right)^{\frac{1}{p}}$$

donc par (5)

$$W_p(\mu_1, \mu_3) \leq W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3).$$

$\square$

3 Problème dual

Nous allons étudier le problème de Monge-Kantorovich en trouvant une autre formulation du coût de transport de Monge-Kantorovich. Nous verrons donc une formulation duale du problème de Monge-Kantorovich qui va apporter les outils principaux pour toute la suite.

3.1 Théorème de dualité de Kantorovich

Les résultats de cette partie proviennent en partie du chapitre 3 de [GSZ18].

Théorème 3.1 (dualité de Kantorovich)**.** *Soit* $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$ *une fonction de coût continue et* μ, ν *des mesures de probabilités sur* $\mathbb{R}^d$ *telles que* $T_c(\mu, \nu) < +\infty$.
Soit $\Phi_c = \left\{(\varphi, \psi) \in L^1(\mu) \times L^1(\nu); \forall x, y \in \mathbb{R}^d,\ \varphi(x) + \psi(y) \leq c(x, y)\right\}$. *Alors*

$$T_c(\mu, \nu) = \sup_{(\varphi,\psi)\in\Phi_c} \left\{ \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) \right\}.$$

Preuve. (Première inégalité) Si $(\varphi, \psi) \in \Phi_c$, alors pour $\pi \in \Pi(\mu, \nu)$,

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} (\varphi(x) + \psi(y)) d\pi(x, y) \leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} c(x, y) d\pi(x, y) = I_c(\pi).$$

Donc en optimisant à gauche et à droite de l'inéquation,

$$\sup_{(\varphi,\psi)\in\Phi_c} \left\{ \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) \right\} \leq \inf_{\pi\in\Pi(\mu,\nu)} I_c(\pi) = T_c(\mu, \nu). \tag{6}$$

□

On va devoir introduire d'abord d'autres outils avant de prouver l'autre sens de l'égalité. On va étudier une condition nécessaire pour qu'un couplage de μ et ν soit optimal.

3.1.1 Condition nécessaire d'optimalité d'un couplage et preuve du théorème de dualité

Définition 3.2 (inf-convolution, sup-convolution et c-convexité)**.** Soit $f : \mathbb{R}^d \longrightarrow \overline{\mathbb{R}}$, on définit respectivement sa sup-convolution et son inf-convolution par :

$$f^c : y \in \mathbb{R}^d \mapsto \sup_{x\in\mathbb{R}^d} \{f(x) - c(x, y)\} \text{ et } f_c : y \mapsto \inf_{x\in\mathbb{R}^d} \{f(x) + c(x, y)\}.$$

On dit que f est c-convexe s'il existe $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ telle que $f = g^c$.

Proposition 3.3. *Une fonction* $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ *est c-convexe si et seulement si* $(f_c)^c = f$. *Dans le cas général,* $(f_c)^c \leq f$.

Preuve. $(f_c)^c(x) = \sup_{y\in\mathbb{R}^d} \inf_{z\in\mathbb{R}^d}\{f(z) + c(z, y) - c(x, y)\}$. Si $z = x$, $f(z) + c(z, y) - c(x, y) = f(x)$, donc directement, $(f_c)^c \leq f$. Supposons f c-convexe. On veut montrer l'inégalité dans l'autre sens. Par hypothèse il existe $g : \mathbb{R}^d \to \overline{\mathbb{R}}$ telle que $f = g^c$. Alors, pour $x \in \mathbb{R}^d$,

$$(f_c)^c(x) = \sup_{y\in\mathbb{R}^d} \inf_{z\in\mathbb{R}^d} \{g^c(z) + c(z, y) - c(x, y)\} = \sup_{y\in\mathbb{R}^d} \inf_{z\in\mathbb{R}^d} \sup_{u\in\mathbb{R}^d} \{g(u) - c(z, u) + c(z, y) - c(x, y)\}.$$

Alors pour $u = y$,

$$(f_c)^c(x) \geq \sup_{y \in \mathbb{R}^d} \inf_{z \in \mathbb{R}^d} \{g(y) - c(x,y)\} = \sup_{y \in \mathbb{R}^d} \{g(y) - c(x,y)\} = g^c(x) = f(x).$$

Donc effectivement quand f est c-convexe, $f = (f_c)^c$. Réciproquement, si $f = (f_c)^c$, f est c-convexe par définition avec $g = f_c$. □

Définition 3.4 (Monotonie cyclique)**.** Soit $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$. Γ est c-cycliquement monotone si

$$\forall p \in \mathbb{N}^*, \forall (x_1, y_1), \ldots, (x_p, y_p) \in \Gamma, \sum_{i=1}^{p} c(x_i, y_i) \leq \sum_{i=1}^{p} c(x_i, y_{i+1}), \text{ en posant } y_{p+1} = y_1.$$

Théorème 3.5 (Condition nécessaire d'optimalité d'un couplage)**.** *Soit $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ continue.*

1. *Soient $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ tels que $T_c(\mu, \nu) < +\infty$. Si un couplage $\pi \in \Pi(\mu, \nu)$ est optimal, alors son support est c-cycliquement monotone.*
2. *Soit $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ un ensemble c-cycliquement monotone. Alors il existe une fonction $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, c-convexe, mesurable, telle que $f < +\infty$ sur $\{x \in \mathbb{R}^d ; \exists y \in \mathbb{R}^d, (x, y) \in \Gamma\}$ et que $\Gamma \subset \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d ; f_c(y) = f(x) + c(x, y)\}$.*

En admettant ce théorème momentanément, nous allons pouvoir terminer de prouver le théorème de dualité de Kantorovich.

Preuve. (Deuxième inégalité) Soit un couplage optimal $\pi^* \in \Pi(\mu, \nu)$ et $\Gamma = \text{Supp}(\pi^*)$. D'après le théorème énoncé, il existe une fonction $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ c-convexe telle que $f_c(y) = f(x) + c(x, y)$ pour π^* presque tout (x, y), et telle que $f < +\infty$ sur $\{x \in \mathbb{R}^d ; \exists y \in \mathbb{R}^d, (x, y) \in \Gamma\}$. Supposons d'abord que c est bornée par M. On va montrer que $f \in L^1(\mu)$ et $f_c \in L^1(\nu)$. Ainsi on pourra séparer les intégrales de cette manière

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi^*(x, y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} (f_c(y) - f(x)) d\pi^*(x, y) = \int_{\mathbb{R}^d} (-f)(x) d\mu(x) + \int_{\mathbb{R}^d} f_c(y) d\nu(y) \quad (7)$$

et par définition, $f_c(y) - f(x) \leq c(x, y)$ pour tout $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ donc $(-f, f_c) \in \Phi_c$ et

$$\int_{\mathbb{R}^d} (-f)(x) d\mu(x) + \int_{\mathbb{R}^d} f_c(y) d\nu(y) \leq \sup_{(\varphi, \psi) \in \Phi_c} \left\{ \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) \right\} \leq T_c(\mu, \nu). \quad (8)$$

La dernière inégalité provenant de (6). En combinant alors (7) et (8), comme π^* est un plan de transport optimal,

$$T_c(\mu, \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi^*(x, y) \leq \sup_{(\varphi, \psi) \in \Phi_c} \left\{ \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y) \right\} \leq T_c(\mu, \nu).$$

On obtient donc bien l'égalité voulue. Il reste à montrer que $f \in L^1(\mu)$ et $f_c \in L^1(\nu)$. Pour $y \in \mathbb{R}^d$,

$$f_c(y) = \inf_{x \in \mathbb{R}^d} \{f(x) + c(x, y)\} \leq \inf_{x \in \mathbb{R}^d} f(x) + M.$$

Donc

$$(f_c)^c(y) = \sup_{x \in \mathbb{R}^d} \{f_c(x) - c(x, y)\} \leq \sup_{x \in \mathbb{R}^d} \{f_c(x)\} + M \leq \inf_{x \in \mathbb{R}^d} f(x) + 2M.$$

Or f étant c-convexe, $(f_c)^c = f$ d'après la proposition (3.3),

$$f \leq \inf_{x \in \mathbb{R}^d} f(x) + 2M,$$

ce qui implique d'une part (comme $f > -\infty$) que $\inf_{x \in \mathbb{R}^d} f(x) > -\infty$. D'autre part f n'est pas identiquement infinie donc $\inf_{x \in \mathbb{R}^d} f(x) < +\infty$, ainsi f est majorée. Donc f est bornée, disons par M_f. Directement, $|f_c| \leq M + M_f$. On sait déjà que f est mesurable, donc f_c l'est également comme sup de fonctions mesurables, donc $f \in L^1(\mu)$ et $f_c \in L^1(\nu)$.

On veut maintenant s'affranchir de l'hypothèse que c est bornée. On ne suppose donc plus que la continuité pour c. Posons

$$c_n : (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \min\{c(x, y), n\} = \begin{cases} c(x, y) \text{ si } c(x, y) \leq n \\ n \text{ sinon} \end{cases}$$

On vérifie que (c_n) est une suite croissante de fonctions continues bornées qui converge simplement vers c.

En appliquant le raisonnement précédent, il existe π_n^* une mesure optimale pour le coût de transport c_n et f_n une fonction mesurable, c_n-convexe et bornée, telle que

$$\int c_n(x, y) d\pi_n^*(x, y) = \int ((f_n)_c(y) - f_n(x)) d\pi_n^*(x, y) = \int_{\mathbb{R}^d} (-f_n)(x) d\mu(x) + \int_{\mathbb{R}^d} (f_n)_c(y) d\nu(y). \quad (9)$$

Par définition, $(-f_n, (f_n)_c) \in \Phi_{c_n}$ et $\Phi_{c_n} \subset \Phi_c$ car

$$\forall (\varphi, \psi) \in \Phi_{c_n}, \forall x, y \in \mathbb{R}^d, \varphi(x) + \psi(y) \leq c_n(x, y) \leq c(x, y).$$

Ainsi, par (9),

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c_n(x, y) d\pi_n^*(x, y) \leq \sup_{(\varphi, \psi) \in \Phi_c} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi d\mu + \int_{\mathbb{R}^d} \psi d\nu \right\}.$$

La suite (c_n) étant croissante, pour $m \leq n$,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c_m(x, y) d\pi_n^*(x, y) \leq \sup_{(\varphi, \psi) \in \Phi_c} \left\{ \int_{\mathbb{R}^d} \varphi d\mu + \int_{\mathbb{R}^d} \psi d\nu \right\}. \quad (10)$$

Or, comme vu en proposition 2.14, $\Pi(\mu, \nu)$ est compact dans l'ensemble des mesures de probabilités sur $\mathbb{R}^d \times \mathbb{R}^d$ pour la topologie de la convergence étroite, ainsi il existe $\pi^* \in \Pi(\mu, \nu)$ telle que π_n^* converge étroitement vers π^*. A m fixé, c_m étant bornée

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c_m(x, y) d\pi_n^*(x, y) \underset{n \to +\infty}{\longrightarrow} \int_{\mathbb{R}^d \times \mathbb{R}^d} c_m(x, y) d\pi^*(x, y)$$

puis par convergence monotone

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c_m(x, y) d\pi^*(x, y) \underset{m \to +\infty}{\longrightarrow} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi^*(x, y).$$

On passe donc à la limite (selon n puis m) dans (10) :

$$T_c(\mu, \nu) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\pi^*(x, y) \leq \sup_{(\varphi, \psi) \in \Phi_c} \left\{ \int_{\mathbb{R}^d} \varphi d\mu + \int_{\mathbb{R}^d} \psi d\nu \right\} \leq T_c(\mu, \nu)$$

donc d'une part, π^* est optimale et d'autre part l'égalité voulue est vérifiée. $\square$

3.1.2 Preuve de la condition nécessaire d'optimalité

Preuve. (1. du théorème 3.5) Prenons un couplage $\pi \in \Pi(\mu, \nu)$ en supposant que son support n'est pas cycliquement monotone. On veut donc montrer que π n'est pas optimal, c'est-à-dire qu'on cherche un autre couplage $\widetilde{\pi}$ tel que $I_c(\pi) > I_c(\widetilde{\pi})$.
Le support n'est pas cycliquement monotone, c'est-à-dire qu'il existe $p \in \mathbb{N}^*$ et $(x_1, y_1), \dots, (x_p, y_p) \in \mathbb{R}^d \times \mathbb{R}^d$, (ici $y_{p+1} = y_1$)

$$\sum_{i=1}^{p} c(x_i, y_i) > \sum_{i=1}^{p} c(x_i, y_{i+1}). \tag{11}$$

L'application $((u_1, v_1), \dots, (u_p, v_p)) \in (\mathbb{R}^d \times \mathbb{R}^d)^p \mapsto \sum_{i=1}^{p} (c(u_i, v_i) - c(u_i, v_{i+1}))$ (avec $v_{p+1} = v_1$) est continue (pour la topologie produit) car c est elle-même continue, donc, d'après (11) est strictement positive au voisinage de $((x_1, y_1), \dots, (x_p, y_p))$. Ainsi il existe $\varepsilon > 0$ tel que pour tout i, il existe des voisinages ouverts U_i de x_i et V_i de y_i tels que

$$\forall((u_1, v_1), \dots, (u_p, v_p)) \in (U_1 \times V_1) \times \cdots \times (U_p \times V_p), \sum_{i=1}^{p} (c(u_i, v_i) - c(u_i, v_{i+1})) > \varepsilon.$$

Remarque 3.6. L'idée ensuite est d'exploiter cet écart d'epsilon pour trouver une mesure un peu plus optimale que π. Concrètement on cherche des variables aléatoires X_i et Y_i à valeurs respectives dans U_i et V_i de sorte que $\mathbb{E}\left[\sum_{i=1}^{p} (c(X_i, Y_i) - c(X_i, Y_{i+1}))\right] > \varepsilon$. Ainsi si γ_i et $\widetilde{\gamma_i}$ désignent les lois respectives de (X_i, Y_i) et (X_i, Y_{i+1}), alors $\widetilde{\pi} := \pi - \sum_{i=1}^{p} (\gamma_i - \widetilde{\gamma_i})$ vérifiera

$$I_c(\widetilde{\pi}) = I_c(\pi) - \int c(x, y) \sum_{i=1}^{p} (d\gamma_i - d\widetilde{\gamma_i})(x, y) = I_c(\pi) - \mathbb{E}\left[\sum_{i=1}^{p} (c(X_i, Y_i) - c(X_i, Y_{i+1}))\right] < I_c(\pi) - \varepsilon.$$

Etant donné que $(x_i, y_i) \in \operatorname{Supp}(\pi)$, alors $m_i := \pi(U_i \times V_i) > 0$ pour tout i. On peut donc définir la mesure (de probabilités) de restriction de π à $U_i \times V_i$

$$\gamma_i : B \in \mathscr{B}(\mathbb{R}^d) \mapsto \frac{1}{m_i} \pi(B \cap (U_i \times V_i)).$$

Soit donc (X_i, Y_i) une variable aléatoire de loi γ_i et on pose $\widetilde{\gamma_i}$ la loi du couple (X_i, Y_{i+1}) (avec bien sûr $Y_{p+1} = Y_1$).
On veut donc poser $\widetilde{\pi} = \pi - \sum_{i=1}^{p} (\gamma_i - \widetilde{\gamma_i})$, et il faut vérifier que $\widetilde{\pi} \in \Pi(\mu, \nu)$, c'est à dire que $\widetilde{\pi}$ est une probabilité de marginales μ et ν. Mais il faut d'abord être certain que c'est une mesure positive, ce qui peut-être réglé ainsi : pour f mesurable positive,

$$\int f(x, y) d\left(\sum_{i=1}^{p} \gamma_i\right)(x, y) = \int f(x, y) \sum_{i=1}^{p} \frac{\mathbf{1}_{U_i \times V_i}(x, y)}{m_i} d\pi(x, y) \leq \frac{p}{\min_{1 \leq i \leq p}\{m_i\}} \int f(x, y) d\pi(x, y).$$

Donc en posant $m := \min_{1\le i\le p}\{m_i\}$, on peut décider d'introduire $\widetilde{\pi} := \pi - \frac{m}{p}\sum_{i=1}^{p}(\gamma_i - \widetilde{\gamma}_i)$ qui est positive. Pour finir en montrant que $\widetilde{\pi} \in \Pi(\mu,\nu)$,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} d\widetilde{\pi}(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} d\pi - \frac{m}{p}\sum_{i=1}^{p}\left(\int_{\mathbb{R}^d} d\widetilde{\gamma}_i - \int_{\mathbb{R}^d} d\gamma_i\right) = 1.$$

Et pour $\psi : \mathbb{R}^d \longrightarrow \mathbb{R}$ mesurable bornée,

$$\int \psi(y)d\widetilde{\pi}(x,y) - \int \psi(y)d\pi(x,y) = -\frac{m}{p}\sum_{i=1}^{p}\left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(y)d\widetilde{\gamma}_i(x,y) - \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(y)d\gamma_i(x,y)\right)$$

$$= -\frac{m}{p}\left(\sum_{i=1}^{p}\mathbb{E}\left[\psi(Y_{i+1})\right] - \sum_{i=1}^{p}\mathbb{E}\left[\psi(Y_i)\right]\right) = 0 \text{ car } Y_{p+1} = Y_1, \text{ donc } \int \psi(y)d\widetilde{\pi}(x,y) = \int_{\mathbb{R}^d} \psi(y)d\nu(y).$$

Et un raisonnement similaire permet d'obtenir que μ est la première marginale de $\widetilde{\pi}$. Ainsi $\pi \in \Pi(\mu,\nu)$ et par le raisonnement expliqué en remarque (3.6),

$$I_c(\widetilde{\pi}) < I_c(\pi) - \frac{m}{p}\varepsilon.$$

Donc π n'est pas optimal. □

Preuve. (2. du théorème (3.5)) Soit $a \in \{x \in \mathbb{R}^d; \exists y \in \mathbb{R}^d, (x,y) \in \Gamma\}$ choisi arbitrairement. On pose

$$f_a : x \in \mathbb{R}^d \mapsto \sup_{p\in\mathbb{N}^*} \sup_{\substack{(x_1,y_1),\dots,(x_p,y_p)\in\Gamma \\ x_p=a, x_0=x}} \left\{\sum_{i=1}^{p}(c(x_i,y_i) - c(x_{i-1},y_i))\right\}.$$

Montrons que f_a est c-convexe. En effet, pour $x \in \mathbb{R}^d$, (les sups sont manipulés avec la convention $\sup\emptyset = -\infty$),

$$f_a(x) = \sup_{p\ge 2} \sup_{\substack{(x_1,y_1),\dots,(x_p,y_p)\in\Gamma \\ x_p=a, x_0=x}} \left\{c(x_1,y_1) - c(x,y_1) + \sum_{i=2}^{p}(c(x_i,y_i) - c(x_{i-1},y_i))\right\}$$

$$= \sup_{(x_1,y_1)\in\Gamma} \left\{-c(x,y_1) + c(x_1,y_1) - \sup_{\substack{p\ge 2,(x_2,y_2),\dots,(x_p,y_p)\in\Gamma \\ x_p=a}} \sum_{i=2}^{p}(c(x_i,y_i) - c(x_{i-1},y_i))\right\}$$

$$= \sup_{y_1\in\mathbb{R}^d} \left\{ \sup_{x_1\in\mathbb{R}^d,(x_1,y_1)\in\Gamma} \left\{c(x_1,y_1) - \sup_{\substack{p\ge 2,(x_2,y_2),\dots,(x_p,y_p)\in\Gamma \\ x_p=a}} \sum_{i=2}^{p}(c(x_i,y_i) - c(x_{i-1},y_i))\right\} - c(x,y_1)\right\}.$$

Donc on voit bien que f_a est une fonction c-convexe en posant

$$g_a : y \in \mathbb{R}^d \mapsto \sup_{x_1\in\mathbb{R}^d,(x_1,y)\in\Gamma} \left\{c(x_1,y) - \sup_{\substack{p\ge 2,(x_2,y_2),\dots,(x_p,y_p)\in\Gamma \\ x_p=a}} \sum_{i=2}^{p}(c(x_i,y_i) - c(x_{i-1},y_i))\right\}$$

On observe bien que $f_a(x) = \sup_{y\in\mathbb{R}^d} \{g_a(y) - c(x,y)\}$. On veut maintenant montrer que

$$\begin{aligned}\Gamma \subset \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d ; (f_a)_c(y) = f_a(x) + c(x,y)\} \\ = \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d ; \forall z \in \mathbb{R}^d, f_a(x) + c(x,y) \leq f_a(z) + c(z,y)\}.\end{aligned}$$

C'est à dire qu'on veut montrer que pour tout $(x,y) \in \Gamma$ et tout $z \in \mathbb{R}^d$,

$$f_a(z) \geq f_a(x) + c(x,y) - c(z,y). \tag{12}$$

$$\begin{aligned} f_a(z) &= \sup_{p\in\mathbb{N}^*} \sup_{\substack{(x_1,y_1),\ldots,(x_p,y_p)\in\Gamma \\ x_p=a, x_0=z}} \left\{ \sum_{i=1}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)) \right\} \\ &\geq \sup_{p\geq 2} \sup_{\substack{(x_1,y_1),\ldots,(x_p,y_p)\in\Gamma \\ (x_1,y_1)=(x,y), x_0=z, x_p=a}} \left\{ \sum_{i=1}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)) \right\} \\ &= c(x,y) - c(z,y) + \sup_{p\geq 2} \sup_{\substack{(x_1,y_1),\ldots,(x_p,y_p)\in\Gamma \\ x_1=x, x_p=a}} \left\{ \sum_{i=2}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)) \right\}. \end{aligned} \tag{13}$$

On aimerait maintenant, pour vérifier (12), montrer que

$$\begin{aligned} &\sup_{p\geq 2} \sup_{\substack{(x_1,y_1),\ldots,(x_p,y_p)\in\Gamma \\ x_1=x, x_p=a}} \left\{ \sum_{i=2}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)) \right\} \\ &\qquad \geq f_a(x) = \sup_{p\in\mathbb{N}^*} \sup_{\substack{(x_1,y_1),\ldots,(x_p,y_p)\in\Gamma \\ x_p=a, x_0=x}} \left\{ \sum_{i=1}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)) \right\} \end{aligned} \tag{14}$$

ce qui prouvera (12) avec l'aide de (13). Or, avec les contraintes respectivement du membre de gauche et du membre de droite, pour $p \geq 3$

$$\sum_{i=2}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)) = c(x_2,y_2) - c(x,y_2) + \sum_{i=3}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)); \tag{15}$$

$$\sum_{i=1}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)) = c(x_1,y_1) - c(x,y_1) + c(x_2,y_2) - c(x_1,y_2) + \sum_{i=3}^{p} (c(x_i,y_i) - c(x_{i-1},y_i)). \tag{16}$$

Donc on voudrait

$$\begin{aligned} c(x_1,y_1) - c(x,y_1) + c(x_2,y_2) - c(x_1,y_2) \leq c(x_2,y_2) - c(x,y_2) \\ \iff c(x_1,y_1) + c(x,y_2) \leq c(x,y_1) + c(x_1,y_2). \end{aligned}$$

Ce qui est vrai en appliquant l'hypothèse de c-cyclique monotonie de Γ au cas $p = 2$, aux couples $((x_1,y_1),(x,y_2))$, donc (15) $\leq$ (16). On prouve de même pour $p = 1,2$, donc en passant au sup, on vérifie bien (14). Il nous reste désormais à prouver que f_a est à valeurs dans $\mathbb{R} \cup \{+\infty\}$ et $f_a < +\infty$ sur l'ensemble $\{x \in \mathbb{R}^d ; \exists y \in \mathbb{R}^d, (x,y) \in \Gamma\}$. Or, pour $y \in \mathbb{R}^d$ tel que $(a,y) \in \Gamma$, pour $p = 1$,

$$f_a(a) \geq c(a,y) - c(a,y) = 0. \tag{17}$$

D'autre part, la quantité à maximiser dans $f_a(a)$ est, sous les contraintes $x_0 = a, x_p = a$,

$$\sum_{i=1}^{p}(c(x_i,y_i)-c(x_{i-1},y_i)) = \sum_{i=1}^{p-1}(c(x_i,y_i)-c(x_i,y_{i+1}))+c(a,y_p)-c(a,y_1) = \sum_{i=1}^{p}(c(x_i,y_i)-c(x_i,y_{i+1})).$$

avec $y_{p+1} = 0$ comme dans la définition de la c-monotonie cyclique. Par cette même définition, cette quantité est toujours négative, donc $f_a(a) \leq 0$, donc par (17), $f_a(a) = 0$. Donc en observant (12) avec $x = a$, on voit bien que $f_a > -\infty$, donc f_a est à valeurs dans $\mathbb{R} \cup \{+\infty\}$. Il ne reste plus qu'à montrer que $f_a < +\infty$ sur l'ensemble $\{x \in \mathbb{R}^d; \exists y \in \mathbb{R}^d, (x,y) \in \Gamma\}$. En remplaçant z par a dans (12), c'est bien ce qu'on observe. $\square$

3.2 Cas du coût quadratique

Les résultats de cette section proviennent en partie du chapitre 2 de [Vil03].

3.2.1 Cadre particulier et reformulation du problème dual

A partir de maintenant c est la fonction $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto ||x-y||^2$. On note $I(\pi) = I_c(\pi)$, $T(\mu,\nu) = T_c(\mu,\nu)$, $\Phi_c = \Phi$. La dualité de Kantorovich (theorème 3.1) énonce qu'en notant

$$J : (\varphi,\psi) \in L^1(\mu) \times L^1(\nu) \mapsto \int_{\mathbb{R}^d} \varphi d\mu + \int_{\mathbb{R}^d} \psi d\nu, \tag{18}$$

alors

$$\inf_{\pi\in\Pi(\mu,\nu)} I_c(\pi) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} ||x-y||^2 d\pi(x,y) = \sup_{(\varphi,\psi)\in\Phi} J(\varphi,\psi).$$

On va chercher à écrire sous une forme différente cette dualité. Tout d'abord, sous l'hypothèse que μ et ν soient de moment d'ordre 2 finis,

$$\begin{aligned}\inf_{\pi\in\Pi(\mu,\nu)} \int ||x-y||^2 d\pi(x,y) &= \inf_{\pi\in\Pi(\mu,\nu)} \left\{ \int ||x||^2 d\mu(x) + \int ||y||^2 d\nu(y) - 2\int (x\cdot y) d\pi(x,y) \right\} \\ &= \int ||x||^2 d\mu(x) + \int ||y||^2 d\nu(y) - 2 \sup_{\pi\in\Pi(\mu,\nu)} \int (x\cdot y) d\pi(x,y). \end{aligned} \tag{19}$$

Soient $(\varphi,\psi) \in \Phi$. C'est-à-dire que pour tous $x, y \in \mathbb{R}^d$,

$$\varphi(x) + \psi(y) \leq ||x-y||^2 \iff x\cdot y \leq \left(\frac{||x||^2 - \varphi(x)}{2}\right) + \left(\frac{||y||^2 - \psi(y)}{2}\right).$$

On pose désormais alors

$$\widetilde{\Phi} = \{(\varphi,\psi) \in L^1(\mu) \times L^1(\nu); \forall x, y \in \mathbb{R}^d, \varphi(x) + \varphi(y) \geq x\cdot y\} \tag{20}$$

et

$$\widetilde{\varphi} : x \in \mathbb{R}^d \mapsto \frac{||x||^2 - \varphi(x)}{2}, \ \widetilde{\psi} : y \in \mathbb{R}^d \mapsto \frac{||y||^2 - \psi(y)}{2}$$

et dans ce cas,

$$J\left(\widetilde{\varphi}, \widetilde{\psi}\right) = \int_{\mathbb{R}^d} \frac{||x||^2}{2} d\mu(x) + \int_{\mathbb{R}^d} \frac{||y||^2}{2} d\nu(y) - \frac{1}{2} J(\varphi,\psi)$$

donc en passant à l'inf, en se servant de (19), en notant $\widetilde{I}_c(\pi) = \int_{\mathbb{R}^d} (x \cdot y) d\pi(x, y)$,

$$\inf_{(\widetilde{\varphi},\widetilde{\psi})\in\widetilde{\Phi}} J\left(\widetilde{\varphi}, \widetilde{\psi}\right) = \int_{\mathbb{R}^d} \frac{||x||^2}{2} d\mu(x) + \int_{\mathbb{R}^d} \frac{||y||^2}{2} d\nu(y) - \frac{1}{2} \sup_{(\varphi,\psi)\in\Phi} J(\varphi, \psi)$$
$$= \sup_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d} (x \cdot y) d\pi(x, y) = \sup_{\pi\in\Pi(\mu,\nu)} \widetilde{I}_c(\pi).$$

3.2.2 Atteinte de la borne inférieure par une fonction convexe et s.c.i

Définition 3.7. Soit $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ non identiquement $+\infty$. La transformée de Legendre ou conjuguée convexe de φ est la fonction $\varphi^* : y \in \mathbb{R}^d \mapsto \sup_{x\in\mathbb{R}^d} \{x \cdot y - \varphi(x)\}$

Remarque 3.8. Ainsi définie, $\forall x, y \in \mathbb{R}^d, \varphi(x) + \varphi^*(y) \geq x \cdot y$.

Proposition 3.9. *Soit $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. Alors les énoncés suivants sont équivalents :*

1. *φ est convexe s.c.i*
2. *Il existe $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ telle que $\varphi = \psi^*$*
3. *$\varphi = (\varphi^*)^*$*

Nous admettrons l'implication 1. $\implies$ 3. qui est le théorème de Fenchel-Moreau dont on trouve une démonstration dans [Che13] (théorème 1.5.7 page 20) et dans [Vil03] (proposition 2.5 page 56).

Preuve. 3. implique 2. directement. Montrons que 2. implique 1., soit donc une fonction $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, montrons que ψ^* est convexe s.c.i. Pour $t \in [0, 1]$ et $x, y \in \mathbb{R}^d$,

$$\psi^*(tx+(1-t)y) = \sup_{z\in\mathbb{R}^d} \{(tx + (1 - t)y) \cdot z - \psi(z)\} = \sup_{z\in\mathbb{R}^d} \{t(x \cdot z - \psi(z)) + (1 - t)(y \cdot z - \psi(z))\}$$
$$\leq \sup_{z_1\in\mathbb{R}^d} \{t(x \cdot z_1 - \psi(z_1))\} + \sup_{z_2\in\mathbb{R}^d} \{(1 - t)(x \cdot z_2 - \psi(z_2))\} = t\psi^*(x) + (1 - t)\psi^*(y)$$

donc ψ^* est effectivement convexe, montrons maintenant qu'elle est s.c.i. On va utiliser le point 1. de la proposition 2.13. Soit $r \in \mathbb{R}$, montrons que $\{\psi^* \leq r\}$ est fermé. Or,

$$\{\psi^* \leq r\} = \left\{y \in \mathbb{R}^d; \sup_{x\in\mathbb{R}^d} (x \cdot y - \psi(x)) \leq r\right\} = \{y \in \mathbb{R}^d; \forall x \in \mathbb{R}^d, \psi(x) + r \geq x \cdot y\}$$
$$= \bigcap_{x\in\mathbb{R}^d} \{y \in \mathbb{R}^d; \psi(x) + r \geq x \cdot y\} = \bigcap_{x\in\mathbb{R}^d} (y \mapsto x \cdot y)^{-1}(] - \infty, \psi(x) + r])$$

qui est une intersection de fermés car la fonction $y \mapsto x \cdot y$ est continue donc c'est un ensemble fermé et ce quelque soit r donc ψ^* est s.c.i par la proposition 2.13.

□

Théorème 3.10. *Soient μ, ν deux mesures de probabilités sur $\mathbb{R}^d$ de carré intégrable. Alors il existe une fonction $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convexe s.c.i telle que*

$$J(\varphi, \varphi^*) = \inf_{\widetilde{\Phi}} J.$$

Lemme 3.11. *Soient μ, ν des mesures de probabilités telles que*

$$A := \int_{\mathbb{R}^d} ||x||^2 d\mu(x) + \int_{\mathbb{R}^d} ||y||^2 d\nu(y) < +\infty.$$

Soit $((\varphi_n, \psi_n))_{n\in\mathbb{N}} \subset \widetilde{\Phi}$ une suite minimisant J sur $\widetilde{\Phi}$, i.e. $J(\varphi_n, \psi_n) \underset{n\to+\infty}{\longrightarrow} \inf_{\widetilde{\Phi}} J$. Alors il existe une suite $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ telle que le couple de fonctions

$$\left(\overline{\varphi_n}, \overline{\psi_n}\right) := (\varphi_n^{**} + a_n, \varphi_n^* - a_n)$$

minimise toujours J sur $\widetilde{\Phi}$ et vérifie

$$\forall x \in \mathbb{R}^d, \forall y \in \mathbb{R}^d, \overline{\varphi_n}(x) + \frac{||x||^2}{2} \geq 0, \quad \overline{\psi_n}(y) + \frac{||y||^2}{2} \geq 0. \tag{21}$$

Preuve. Avant tout, prouvons que φ_n^* n'est pas identiquement infinie. Pour $n \in \mathbb{N}$, $(\varphi_n, \psi_n) \in \widetilde{\Phi}$ donc pour tous $x, y \in \mathbb{R}^d$, $\varphi_n(x) + \psi_n(y) \geq x \cdot y$. Or ψ_n n'est pas identiquement $+\infty$ (car intégrable) donc il existe $y_0 \in \mathbb{R}^d$ tel que $\psi_n(y_0) < +\infty$. Pour tout $x \in \mathbb{R}^d$, $\varphi_n(x) \geq x \cdot y_0 - \psi_n(y_0)$, ainsi

$$\varphi_n^*(y_0) = \sup_{x\in\mathbb{R}^d} \{x \cdot y_0 - \varphi_n(x)\} \leq \psi_n(y_0).$$

Donc φ_n^* n'est pas identiquement $+\infty$, et on peut poser

$$a_n := \inf_{y\in\mathbb{R}^d} \left\{\varphi_n^*(y) + \frac{||y||^2}{2}\right\} < +\infty.$$

On peut poser comme énoncé,

$$\left(\overline{\varphi_n}, \overline{\psi_n}\right) := (\varphi_n^{**} + a_n, \varphi_n^* - a_n)$$

Avec donc $\overline{\varphi_n} = \left(\overline{\psi_n}\right)^*$, donc pour $x \in \mathbb{R}^d$,

$$\begin{aligned}\overline{\varphi_n}(x) + \frac{||x||^2}{2} = \left(\overline{\psi_n}\right)^*(x) + \frac{||x||^2}{2} &= \sup_{y\in\mathbb{R}^d} \left\{x \cdot y - \overline{\psi_n}(y) + \frac{||x||^2}{2}\right\} \geq \sup_{y\in\mathbb{R}^d} \left\{-\overline{\psi_n}(y) - \frac{||y||^2}{2}\right\} \\ &= -\inf_{y\in\mathbb{R}^d} \left\{\overline{\psi_n}(y) + \frac{||y||^2}{2}\right\} = -\inf_{y\in\mathbb{R}^d} \left\{\varphi_n^*(y) + \frac{||y||^2}{2} - a_n\right\} = -a_n + a_n = 0.\end{aligned}$$

On vient donc de prouver (21) (pour $\overline{\psi_n}$, c'est par $\inf_{y\in\mathbb{R}^d} \left\{\overline{\psi_n}(y) + \frac{||y||^2}{2}\right\} = 0$). On veut vérifier que $\left(\left(\overline{\varphi_n}, \overline{\psi_n}\right)\right)_{n\in\mathbb{N}}$ reste une suite minimisant J sur $\widetilde{\Phi}$. Montrons déjà que $\left(\overline{\varphi_n}, \overline{\psi_n}\right) \in \widetilde{\Phi}$, et d'abord que ces fonctions sont intégrables. Comme $(\varphi_n, \psi_n) \in \widetilde{\Phi}$, alors

$$\forall x, y \in \mathbb{R}^d, \varphi_n(x) + \psi_n(y) \geq x \cdot y \implies \forall y \in \mathbb{R}^d, \psi_n(y) \geq \sup_{x\in\mathbb{R}^d} \{x \cdot y - \varphi_n(x)\} = \varphi_n^*(y). \tag{22}$$

On en déduit (croissance de l'intégrale) que

$$J(\varphi_n, \psi_n) \geq J(\varphi_n, \varphi_n^*).$$

Et d'après la remarque 3.8, alors on peut effectuer le même raisonnement qu'en (22) pour obtenir :

$$\forall x, y \in \mathbb{R}^d, \varphi_n^*(x) + \varphi_n^{**}(y) \geq x \cdot y \implies \forall x \in \mathbb{R}^d, \varphi_n(x) \geq \sup_{y\in\mathbb{R}^d} \{x \cdot y - \varphi_n^*(y)\} = \varphi_n^{**}(x) \tag{23}$$

donc

$$+\infty > J(\varphi_n, \psi_n) \geq J(\varphi_n, \varphi_n^*) \geq J(\varphi_n^{**}, \varphi_n^*) = J(\varphi_n^{**} - a_n, \varphi_n^* + a_n) = J\left(\overline{\varphi_n}, \overline{\psi_n}\right). \tag{24}$$

L'avant dernière égalité étant obtenue par linéarité. En particulier,

$$+\infty > J\left(\overline{\varphi_n}, \overline{\psi_n}\right) + \frac{A}{2} = \int_{\mathbb{R}^d} \left(\overline{\varphi_n}(x) + \frac{||x||^2}{2}\right) d\mu(x) + \int_{\mathbb{R}^d} \left(\overline{\psi_n}(y) + \frac{||y||^2}{2}\right) d\nu(y),$$

Et d'après (21), comme les intégrandes sont positives, elles sont intégrables. Or, pour $x \in \mathbb{R}^d$, $\overline{\varphi_n}(x) = (\overline{\varphi_n} + \frac{||x||^2}{2}) - \frac{||x||^2}{2}$, somme de deux quantités intégrables, donc $\overline{\varphi_n}$ est intégrable, de même pour $\overline{\psi_n}$. Par leur définition et ce qui est évoqué en (23), $\overline{\varphi}_n(x) + \overline{\psi}_n(y) \geq x \cdot y$ pour $x, y \in \mathbb{R}^d$, donc on a bien vérifié que $\left(\overline{\varphi_n}, \overline{\psi_n}\right) \in \widetilde{\Phi}$. Par l'inégalité (24) et l'hypothèse de minimisation sur (φ_n, ψ_n), on vérifie bien que $\left(\left(\overline{\varphi_n}, \overline{\psi_n}\right)\right)_{n\in\mathbb{N}}$ minimise J sur $\widetilde{\Phi}$. □

3.2.3 Preuve du théorème 3.10

On va maintenant pouvoir prouver le théorème 3.10.. L'idée de la preuve est, par un argument de compacité, d'extraire d'une suite $((\varphi_n, \psi_n))_n$ minimisant J sur $\widetilde{\Phi}$, une sous-suite $((\varphi_{u(n)}, \psi_{v(n)})_n$ dont les composantes sont faiblement convergentes respectivement dans $L^1(\mu)$,vers φ et dans $L^1(\nu)$, vers ψ. J étant la somme de deux formes linéaires sur $L^1(\mu)$ et $L^1(\nu)$, $J(\varphi_{u(n)}, \psi_{v(n)}) \longrightarrow J(\varphi, \psi)$. L'aspect annoncé de convexité, de semi-continuité et de conjugaison s'en déduira. Le théorème suivant, admis (voir [DP40], théorème 3.2.1), est précisément le théorème à appliquer à la situation :

Théorème 3.12 (Dunford-Pettis)**.** *Soit $(X, \mathscr{T}, P)$ un espace probabilisé et $(f_n)_n$ une suite bornée dans $L^1(P)$. On peut en extraire une sous-suite faiblement convergente dans $L^1(P)$ si et seulement si $(f_n)_n$ est uniformément intégrable, i.e.*

$$\forall \varepsilon > 0, \exists \delta > 0, \forall B \in \mathscr{T}, P(B) \leq \delta \implies \sup_{n\in\mathbb{N}} \int_B |f_n| dP < \varepsilon, \tag{25}$$

Preuve. (du théorème 3.10) Soit $((\varphi_n, \psi_n))_{n\in\mathbb{N}}$ une suite de fonctions minimisant J sur $\widetilde{\Phi}$, qui peut donc vérifier les hypothèses énoncées dans le lemme 3.11 :

$$\forall x \in \mathbb{R}^d, \forall y \in \mathbb{R}^d, \varphi_n(x) + \frac{||x||^2}{2} \geq 0, \quad \psi_n(y) + \frac{||y||^2}{2} \geq 0. \tag{26}$$

Pour respecter les hypothèses du théorème 3.12, on préfèrera manipuler des fonction tronquées, forcées à être bornées, puis ensuite rattraper la troncature par convergence monotone. Le procédé de troncature sera le même que dans la preuve du théorème 3.1 : pour $k \in \mathbb{N}$ on pose pour $x \in \mathbb{R}^d$, $\varphi_n^{(k)}(x) = \min\left\{\varphi_n(x) + \frac{||x||^2}{2}, k\right\} - \frac{||x||^2}{2}$, c'est-à-dire que

$$\varphi_n^{(k)}(x) + \frac{||x||^2}{2} = \begin{cases} \varphi_n(x) + \dfrac{||x||^2}{2} \text{ si } \varphi_n(x) + \dfrac{||x||^2}{2} \leq k \\ k \text{ sinon} \end{cases}.$$

On pose de même pour $y \in \mathbb{R}^d$ $\psi_n^{(k)}(y) = \min\left\{\psi_n(y) + \frac{||y||^2}{2}, k\right\} - \frac{||y||^2}{2}$. On vérifie tout de suite que

$$\forall x \in \mathbb{R}^d, 0 \leq \varphi_n^{(k)}(x) + \frac{||x||^2}{2} \leq k, \text{ et } \forall y \in \mathbb{R}^d, 0 \leq \psi_n^{(k)}(y) + \frac{||y||^2}{2} \leq k. \tag{27}$$

On vérifie aussi qu'à n fixé, $(\varphi_n^{(k)})_k$ et $(\psi_n^{(k)})_k$ sont des suites croissantes et majorées respectivement par φ_n et ψ_n, ce qui permet d'énoncer un fait utile pour plus tard :

$$J\left(\varphi_n^{(k)}, \psi_n^{(k)}\right) \leq J(\varphi_n, \psi_n). \tag{28}$$

On va employer le théorème 3.12 sur $\left(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \mu\right)$ avec la suite de fonctions $(\varphi_n^{(k)} + \frac{||\cdot||^2}{2})_n$, qui est d'après (27) une suite bornée dans $L^1(\mu)$ de fonctions intégrables. Ensuite, montrons l'équi-intégrabilité. Soit $\varepsilon > 0$ fixé, alors pour $B \in \mathscr{B}(\mathbb{R}^d)$, pour $n \in \mathbb{N}$,

$$\int_B \left|\varphi_n^{(k)}(x) + \frac{||x||^2}{2}\right| d\mu(x) \leq \int_B k d\mu \leq k\mu(B),$$

donc en posant $\delta = \varepsilon/k$, on vérifie bien la définition de l'équi-intégrabilité. Donc par le théorème de Dunford-Pettis, il existe une extraction $(u_k(n))_n$ et $\widetilde{\varphi}^{(k)}$ une fonction de $L^1(\mu)$ telle que

$$\varphi_{u_k(n)}^{(k)} + \frac{||\cdot||^2}{2} \underset{n\to+\infty}{\longrightarrow} \widetilde{\varphi}^{(k)} \text{ faiblement dans } L^1(\mu).$$

On pose $\varphi^{(k)} = \widetilde{\varphi}^{(k)} - \frac{||\cdot||^2}{2}$. De même, il existe $(v_k(n))_n$ une extraction et une fonction $\psi^{(k)}$ telle que

$$\varphi_{u_k(n)}^{(k)} \underset{n\to+\infty}{\longrightarrow} \varphi^{(k)} \text{ faiblement dans } L^1(\mu) \text{ et } \psi_{v_k(n)}^{(k)} \underset{n\to+\infty}{\longrightarrow} \psi^{(k)} \text{ faiblement dans } L^1(\nu).$$

Par extraction diagonale, il existe $(u(n))_n$ (resp. $(v(n))_n$) une extraction telle que quelque soit $k \in \mathbb{N}$

$$\varphi_{u(n)}^{(k)} \underset{n\to+\infty}{\longrightarrow} \varphi^{(k)} \text{ faiblement dans } L^1(\mu) \left(\text{resp. } \psi_{v(n)}^{(k)} \underset{n\to+\infty}{\longrightarrow} \psi^{(k)} \text{ faiblement dans } L^1(\nu)\right).$$

Par le passage à la limite permis par les convergences faibles et (28),

$$J\left(\varphi^{(k)}, \psi^{(k)}\right) = \lim_{n\to+\infty} J\left(\varphi_{u(n)}^{(k)}, \psi_{v(n)}^{(k)}\right) \leq \lim_{n\to+\infty} J(\varphi_n, \psi_n) = \inf_{\widetilde{\Phi}} J. \tag{29}$$

Montrons que la convergence faible des suites $(\varphi_{u(n)})^{(k)})_n$ et $(\psi_{v(n)}^{(k)})_n$ conserve leur croissance. Par exemple pour $(\varphi_{u(n)})^{(k)})_n$, le dual topologique de $L^1(\mu)$ étant $L^\infty(d\mu)$, la convergence faible dans $L^1(\mu)$ s'écrit :

$$\forall f \in L^\infty(d\mu), \int_{\mathbb{R}^d} f\varphi_{u(n)}^{(k)} d\mu \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} f\varphi^{(k)} d\mu$$

donc la convergence suivante est justifiée :

$$\int_{\mathbb{R}^d} \mathbf{1}_{\varphi^{(k+1)} < \varphi^{(k)}} \times \left(\varphi_{u(n)}^{(k+1)} - \varphi_{u(n)}^{(k)}\right) d\mu \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} \mathbf{1}_{\varphi^{(k+1)} < \varphi^{(k)}} \times \left(\varphi^{(k+1)} - \varphi^{(k)}\right) d\mu \leq 0. \tag{30}$$

Or, la suite $(\varphi_{u(n)}^{(k)})_k$ étant croissante, la quantité à gauche est positive, donc la quantité à droite est à la fois positive et négative : elle est nulle. Or l'intégrande est de signe constant donc est nulle μ-presque partout : $\mathbf{1}_{\varphi^{(k+1)} < \varphi^{(k)}} \times \left(\varphi^{(k+1)} - \varphi^{(k)}\right) = 0$ μ-p.p, ce qui implique directement que $\varphi^{(k+1)} \geq \varphi^{(k)}$ μ-p.p. Une précision est ici nécessaire : la croissance p.p dépend à chaque fois de k, elle s'écrit ici :

$$\forall k \in \mathbb{N}, \exists N_k \subset \mathscr{B}(\mathbb{R}^d), \mu(N_k) = 0 \text{ et } \left\{\varphi^{(k)} > \varphi^{(k+1)}\right\} \subset N_k$$

ainsi en posant $N = \bigcup_{k=0}^{+\infty} N_k$,

$$\bigcup_{k=0}^{+\infty} \left\{ \varphi^{(k)} > \varphi^{(k+1)} \right\} \subset N \text{ et } \mu(N) \leq \sum_{k=0}^{+\infty} \mu(N_k) = 0$$

ce qui prouve bien que $(\varphi^{(k)})_k$ est une suite croissante μ-p.p. La suite de fonctions $(\psi^{(k)})_k$ est de même une suite croissante ν-p.p dans le même sens. Ainsi il existe p.p de fonctions limites φ et ψ. On peut appliquer le théorème de convergence monotone aux suites de fonctions $(\varphi^{(k)} + \frac{||\cdot||^2}{2})_k$ et $(\psi^{(k)} + \frac{||\cdot||^2}{2})_k$, en effet ce sont des suites croissantes p.p et positives p.p (par exactement le même raisonnement que pour la croissance p.p appliqué aux inégalités de (27)). Ainsi,

$$\int_{\mathbb{R}^d} \left(\varphi^{(k)}(x) + \frac{||x||^2}{2} \right) d\mu(x) + \int_{\mathbb{R}^d} \left(\psi^{(k)}(y) + \frac{||y||^2}{2} \right) d\nu(y) \underset{k \to +\infty}{\longrightarrow} \int_{\mathbb{R}^d} \left(\varphi(x) + \frac{||x||^2}{2} \right) d\mu(x) + \int_{\mathbb{R}^d} \left(\psi(y) + \frac{||y||^2}{2} \right) d\nu(y) \quad (31)$$

et par (29), quelque soit $k \in \mathbb{N}$,

$$\int_{\mathbb{R}^d} \left(\varphi^{(k)}(x) + \frac{||x||^2}{2} \right) d\mu(x) + \int_{\mathbb{R}^d} \left(\psi^{(k)}(y) + \frac{||y||^2}{2} \right) d\nu(y) \leq \inf_{\widetilde{\Phi}} J + A$$

donc avec le passage à la limite (31),

$$\int_{\mathbb{R}^d} \left(\varphi(x) + \frac{||x||^2}{2} \right) d\mu(x) + \int_{\mathbb{R}^d} \left(\psi(y) + \frac{||y||^2}{2} \right) d\nu(y) \leq \inf_{\widetilde{\Phi}} J + A < +\infty$$

c'est-à-dire que $\varphi + \frac{||\cdot||^2}{2}$ et $\psi + \frac{||\cdot||^2}{2}$ sont intégrables (respectivement selon μ et ν) car positives, ce qui implique par l'intégrabilité de $\frac{||\cdot||^2}{2}$ selon μ et ν, que φ et ψ sont intégrables (respectivement selon μ et ν) comme somme de fonctions intégrables (par exemple $\varphi = (\varphi - \frac{||\cdot||^2}{2}) + \frac{||\cdot||^2}{2}$). Ce raisonnement permet de transformer (31) en

$$J\left(\varphi^{(k)}, \psi^{(k)}\right) = \int_{\mathbb{R}^d} \varphi^{(k)} d\mu + \int_{\mathbb{R}^d} \psi^{(k)} d\nu \underset{k \to +\infty}{\longrightarrow} \int_{\mathbb{R}^d} \varphi d\mu + \int_{\mathbb{R}^d} \psi d\nu = J(\varphi, \psi).$$

En passant alors à la limite l'inégalité (29), on obtient

$$J(\varphi, \psi) \leq \inf_{\widetilde{\Phi}} J. \quad (32)$$

Pour montrer que $(\varphi, \psi) \in \widetilde{\Phi}$, il ne reste plus que l'inégalité à montrer. Or,

$$\begin{aligned}
\varphi^{(k)}_{u(n)}(x) + \psi^{(k)}_{v(n)}(y) + \frac{||x||^2}{2} + \frac{||y||^2}{2} &= \min\left\{ \varphi_{u(n)}(x) + \frac{||x||^2}{2}, k \right\} + \min\left\{ \psi_{v(n)}(y) + \frac{||y||^2}{2}, k \right\} \\
&= \begin{cases} \varphi_{u(n)}(x) + \frac{||x||^2}{2} + \psi_{v(n)}(y) + \frac{||y||^2}{2} \\ \text{ou } \varphi_{u(n)}(x) + \frac{||x||^2}{2} + k \\ \text{ou } \psi_{v(n)}(y) + \frac{||y||^2}{2} + k \\ \text{ou } 2k \end{cases} \geq \min\left\{ x \cdot y + \frac{||x||^2}{2} + \frac{||y||^2}{2}, k \right\},
\end{aligned}$$

on a utilisé le fait que $(\varphi_{u(n)}, \psi_{v(n)}) \in \widetilde{\Phi}$ et que $\varphi_{u(n)}(x) + \frac{||x||^2}{2}, \psi_{v(n)}(y) + \frac{||y||^2}{2} \geq 0$. Après passage à la limite en n, justifiée avec exactement le même procédé utilisé en (30) pour prouver la croissance p.p de la suite $(\varphi^{(k)})_k$, pour μ-presque tout $x \in \mathbb{R}^d$ et ν-presque tout $y \in \mathbb{R}^d$, pour tout $k \in \mathbb{N}$,

$$\varphi^{(k)}(x) + \psi^{(k)}(y) \geq \min\left\{x \cdot y + \frac{||x||^2}{2} + \frac{||y||^2}{2}, k\right\} - \frac{||x||^2}{2} - \frac{||y||^2}{2}.$$

En passant à la limite en k, on obtient, pour μ-presque tout x, pour ν-presque tout y,

$$\varphi(x) + \psi(y) \geq x \cdot y.$$

φ et ψ étant définies p.p, on choisit de les définir sur les ensembles négligeables considérés (pour μ et ν) par $\varphi(x) = ||x||^2/2$ et $\psi(y) = ||y||^2/2$ de telle sorte que

$$\forall x, y \in \mathbb{R}^d, \varphi(x) + \psi(y) \geq x \cdot y.$$

Donc $(\varphi, \psi) \in \widetilde{\Phi}$, c'est pourquoi en utilisant l'inégalité (32), c'est un couple minimisateur de J sur $\widetilde{\Phi}$:

$$J(\varphi, \psi) = \inf_{\widetilde{\Phi}} J. \tag{33}$$

Pour terminer, on va effectuer un raisonnement déjà utilisé :

$$\forall x, y \in \mathbb{R}^d, \varphi(x) + \psi(y) \geq x \cdot y \implies \forall y \in \mathbb{R}^d, \psi(y) \geq \sup_{x \in \mathbb{R}^d} \{x \cdot y - \varphi(x)\} = \varphi^*(y). \tag{34}$$

Et de même,

$$\forall x, y \in \mathbb{R}^d, \varphi(x) + \varphi^*(y) \geq x \cdot y \implies \forall x \in \mathbb{R}^d, \varphi(x) \geq \sup_{y \in \mathbb{R}^d} \{x \cdot y - \varphi^*(y)\} = \varphi^{**}(x). \tag{35}$$

Donc, en combinant ces deux inégalités et (33)

$$\inf_{\widetilde{\Phi}} J = J(\varphi, \psi) \geq J(\varphi^{**}, \varphi^*) \tag{36}$$

Il reste à montrer que $(\varphi^{**}, \varphi^*) \in \widetilde{\Phi}$. L'inégalité provient de la remarque 3.8 et en convoquant (34),

$$\forall y \in \mathbb{R}^d, \psi(y) \geq \varphi^*(y) = \sup_{y \in \mathbb{R}^d} \{x \cdot y - \varphi(y)\} \geq 0 \cdot y - \varphi(0) = -\varphi(0).$$

Ce qui constitue un encadrement de φ^* par des fonctions intégrables selon ν (ν étant une mesure de probabilités) donc $\varphi^* \in L^1(\nu)$. De même, $\varphi^{**} \in L^1(\mu)$ en convoquant l'inégalité (35) et une minoration par $-\varphi^*(0)$, ainsi Donc (36) est en fait une égalité ce qui conclut la preuve en convoquant la proposition 3.9. $\square$

4 Potentiels de transport optimal

4.1 Théorème de Brenier

Les résultats de cette section proviennent en partie du chapitre 5 de [GSZ18]. Nous nous plaçons toujours dans le même cadre : les mesures sont définies sur $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ et la fonction de coût est la fonction $c : (x,y) \in \mathbb{R}^d \mapsto ||x-y||^2$. Soient μ, ν deux mesures de probabilités sur $\mathbb{R}^d$ qui possèdent un moment d'ordre 2 et introduisons π^* un plan de transport optimal. Par le théorème 3.10, π^* est un plan de transport optimal si et seulement si il existe une fonction convexe s.c.i $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ telle que

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(x) + \varphi^*(y)) d\pi^*(x,y) = J(\varphi, \varphi^*) = \inf_{\widetilde{\Phi}} J = \int_{\mathbb{R}^d \times \mathbb{R}^d} (x \cdot y) d\pi^*(x,y). \tag{37}$$

Or par définition de φ, $\varphi(x) + \varphi^*(y) \geq x \cdot y$ pour tout $(x,y) \in \mathbb{R}^d$, donc

$$\begin{aligned}
&\pi^* \text{ est un plan de transport optimal si et seulement si il existe une fonction convexe s.c.i} \\
&\qquad \varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\} \text{ telle que pour } \pi^*\text{-presque tout } (x,y),\ \varphi(x) + \varphi^*(y) - x \cdot y = 0 \\
&\qquad\qquad \iff \forall z \in \mathbb{R}^d, \varphi(z) \geq \varphi(x) + (z-x) \cdot y \iff y \in \mathrm{SD}\varphi(x)
\end{aligned} \tag{38}$$

où l'on introduit $\mathrm{SD}\varphi(x) := \{y \in \mathbb{R}^d ; \forall z \in \mathbb{R}^d, \varphi(z) \geq \varphi(x) + (z-x) \cdot y\}$ le sous-différentiel de φ en x. Notons λ la mesure de Lebesgue sur $\mathbb{R}^d$. Nous devons énoncer pour la suite le lemme suivant dont les points non démontrés le sont dans [GSZ18] (théorème 5.2) :

Lemme 4.1. *Soit* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *convexe. On note* $\mathrm{dom}(\varphi) = \{\varphi < +\infty\}$ *et* $\mathrm{domdiff}(\varphi) = \{x \in \text{intérieur}(\mathrm{dom}(\varphi)) ; \varphi \textit{ est différentiable en } x\}$*. Alors,*

1. $\lambda(\partial \mathrm{dom}(\varphi)) = 0$
2. $\lambda(\mathrm{dom}(\varphi) \setminus \mathrm{domdiff}(\varphi)) = 0$
3. *pour tous* $x \in \mathrm{domdiff}(\varphi), z \in \mathrm{dom}(\varphi)$, $\varphi(z) \geq \varphi(x) + (z-x) \cdot \nabla\varphi(x)$
4. *si* φ *vérifie* (38)*, alors pour tout* $x \in \mathrm{domdiff}(\varphi), SD\varphi(x) = \{\nabla\varphi(x)\}$.

Preuve. On va seulement prouver l'alinéa 4. D'après (38), si $x \in \mathrm{domdiff}(\varphi)$ et $y \in \mathrm{SD}\varphi(x)$, alors pour $h \in \mathbb{R}^d$

$$\nabla\varphi(x) \cdot h = \lim_{t \to 0} \frac{\varphi(x+th) - \varphi(x)}{t} \geq h \cdot y \iff (\nabla\varphi(x) - y) \cdot h \geq 0$$

donc quelque soit $h \in \mathbb{R}^d$ (en regardant en $-h$ par exemple), $(\nabla\varphi(x) - y) \cdot h = 0$ puis en prenant $h = \nabla\varphi(x) - y$, $\nabla\varphi(x) = y$. On a donc prouvé que si $x \in \mathrm{domdiff}(\varphi)$, alors $\mathrm{SD}\varphi(x) \subset \{\nabla\varphi(x)\}$. D'après le troisième alinéa, en retour, si $x \in \mathrm{domdiff}(\varphi)$ et $y = \nabla(x)$, alors $y \in \mathrm{SD}\varphi(x)$, c'est-à-dire que $\{\nabla\varphi(x)\} \subset \mathrm{SD}\varphi(x)$. ☐

On peut énoncer le théorème de Brenier, qui découle de ce propos :

Théorème 4.2 (Brenier)**.** *Soient* μ,ν *deux mesures de probabilité sur* $\mathbb{R}^d$*. On suppose que* μ *est absolument continue par rapport à la mesure de Lebesgue sur* $\mathbb{R}^d$ *et que* $T(\mu,\nu) < +\infty$*. Alors il existe un unique plan de transport optimal,* π^**, qui est déterministe, c'est-à-dire qu'il existe une application de transport* T^* *de* μ *sur* ν *telle que* $\pi^*(dxdy) = \mu(dx)\delta_{T^*(x)}(dy)$*. Cette application est de la forme* $T^* = \nabla\varphi$ *où* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *est une fonction convexe* μ*-presque sûrement finie, qu'on appelle* ***potentiel de transport optimal****. De plus, si* $(T')^*$ *est une autre application de transport optimale entre* μ *et* ν*, alors* $T^* = (T')^*$ μ*-presque sûrement.*

Remarque 4.3. Ainsi, pour un plan de transport optimal π^* (donc unique dans le cadre du théorème), un potentiel de transport optimal désigne à la fois une fonction convexe s.c.i $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ telle que $\nabla\varphi$ est une application de transport et de manière équivalente, telle que $J(\varphi, \varphi^*) = \min_{\widetilde{\Phi}} J$.

Preuve. Soit $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, fonction convexe s.c.i telle que $J(\varphi, \varphi^*) = \min_{\widetilde{\Phi}} J$. μ étant absolument continue par rapport à la mesure de Lebesgue, en utilisant l'alinéa 2. du lemme 4.1, $\mu(\text{dom}(\varphi)) = \mu(\text{dom}(\varphi) - \text{domdiff}(\varphi)) + \mu(\text{domdiff}(\varphi)) = \mu(\text{domdiff}(\varphi))$. De ce qu'on a étudié dans le théorème 3.10, on sait que $\varphi \in L^1(\mu)$, donc φ est finie μ-presque sûrement, c'est-à-dire que $\mu(\text{dom}(\varphi)) = 1$, donc $\mu(\text{domdiff}(\varphi)) = 1$. φ est donc différentiable μ-presque partout, et dans ce cas, par l'alinéa 4 du lemme 4.1, pour μ-presque tout $x \in \mathbb{R}^d$, $\text{SD}\varphi(x) = \{\nabla\varphi(x)\}$. Par (38), pour π^*-presque tout $(x, y) \in \mathbb{R}^d$, $y \in \text{SD}\varphi(x)$ donc μ étant une marginale de π^*, pour π^*-presque tout (x, y), $y = \nabla\varphi(x)$. Ainsi pour toute fonction test $h : \mathbb{R}^d \to \mathbb{R}$ continue bornée,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} h(x, y) d\pi^*(x, y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} h(x, \nabla\varphi(x)) d\pi^*(x, y) = \int_{\mathbb{R}^d} h(x, \nabla\varphi(x)) d\mu(x)$$

ce qui veut dire que $\pi^*(dxdy) = \mu(dx)\delta_{\nabla\varphi(x)}(dy)$, donc π^* est une plan de transport optimal déterministe, d'application de transport $\nabla\varphi$. Montrons désormais l'unicité du plan de transport optimal. On suppose qu'il y a deux plans de transport optimaux π_1^* et π_2^*. D'après le travail effecuté, il existe φ_1 et φ_2 qui vérifient les hypothèses du théorème, on veut alors prouver que $\nabla\varphi_1 = \nabla\varphi_2$. φ_1 et φ_2 vérifient tous deux (37) donc

$$\begin{aligned} \int (\varphi_2(x) + \varphi_2^*(y)) d\pi_1^*(x, y) = \int \varphi_2 d\mu + \int \varphi_2^* d\nu = \inf_{\widetilde{\Phi}} J = \int \varphi_1 d\mu + \int \varphi_1^* d\nu \\ = \int (\varphi_1(x) + \varphi_1^*(y)) d\pi_1^*(x, y) = \int (x \cdot y) d\pi_1^*(x, y). \end{aligned} \tag{39}$$

et on a vu que pour π_1^*-presque tout (x, y), $y = \nabla\varphi_1(x)$, donc (39) devient

$$\int (\varphi_2(x) + \varphi_2^*(\nabla\varphi_1(x))) d\mu(x) = \int (x \cdot \nabla\varphi_1(x)) d\mu(x).$$

Or, par définition de φ_2^*, l'intégrande de droite est plus grande que celle de gauche, donc par le même raisonnement qu'en (38),

$$\text{pour } \mu\text{-presque tout } x \in \mathbb{R}^d, \nabla\varphi_1(x) \in \partial\varphi_2(x) \tag{40}$$

mais pour μ-presque tout $x \in \mathbb{R}^d$, $\partial\varphi_2(x) = \{\nabla\varphi_2(x)\}$ (alinéa 4 de 4.1) car φ_2 est différentiable μ-presque partout (déjà vu), donc (40) devient

$$\text{pour } \mu\text{-presque tout } x \in \mathbb{R}^d, \nabla\varphi_1(x) = \nabla\varphi_2(x)$$

ce qui prouve bien l'unicité μ-presque partout annoncée. □

On peut ajouter une dernière proposition

Proposition 4.4. *Soient μ,ν deux mesures de probabilité sur $\mathbb{R}^d$. On suppose que μ est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ et que $T(\mu, \nu) < +\infty$. Alors si φ est un potentiel de transport optimal de μ à ν, pour μ-presque tout $x \in \mathbb{R}^d$, $\nabla\varphi^* \circ \nabla\varphi(x) = x$ et pour ν-presque tout $y \in \mathbb{R}^d$, $\nabla\varphi \circ \nabla\varphi^*(y) = y$.*

Preuve. Soit π^* est un plan de transport optimal. Alors on a vu que pour π^*-presque tout (x, y), $y = \nabla\varphi(x)$. Le même raisonnement que dans le début de la preuve du théorème 4.2, comme φ^* est finie ν-presque partout, alors φ^* est différentiable presque partout et pour π^*-presque tout (x, y), $x = \nabla\varphi^*(y) = \nabla\varphi^*(\varphi(x))$. μ étant la première marginale de π^* donc pour μ-presque tout x, $\nabla\varphi^*(\varphi(x))$. De même pour l'autre égalité par symétrie. $\square$

4.2 Stabilité des potentiels de transport optimal

Bien que dans le cadre du théorème 4.2 il y a unicité de l'application de transport optimal, il n'y a pas unicité du potentiel de transport optimal. En effet si φ_1 en est un alors pour $C \in \mathbb{R}$, $J((\varphi + C), (\varphi + C)^*) = J(\varphi + C, \varphi^* - C) = J(\varphi, \varphi^*)$, la dernière quantité minimisant J donc par le raisonnement de la section précédente $(\varphi + C, (\varphi + C)^*)$ est un potentiel de transport optimal. On peut se demander si tous les potentiels de transport optimal sont sous cette forme. On admet le lemme suivant qui est démontré dans [BL19] (lemme 2.1) :

Lemme 4.5. *Soit $A \subset \mathbb{R}^d$ un convexe non vide et deux fonctions convexes $\varphi_1, \varphi_2 : A \to \mathbb{R}$ telles que pour λ-presque tout $x \in A$, $\nabla\varphi_1(x) = \nabla\varphi_2(x)$. Alors il existe une constante $C \in \mathbb{R}$ telle que $\varphi_1(x) = \varphi_2(x) + C$ pour tout $x \in A$.*

Qui permet ensuite de déduire qu'il y a bien unicité à la constante près du potentiel de transport optimal en ajoutant une hypothèse précise de régularité près. Pour se rapprocher des idées qui se développeront au fur et à mesure, μ et ν seront notées P et Q.

Théorème 4.6. *Soient P et Q deux mesures de probabilités sur $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ de moments d'ordre deux finis. On suppose que P est à support convexe, absolument continue par rapport à λ et que sa densité est strictement positive à l'intérieur de son support. Alors si φ_1 et φ_2 sont deux potentiels de transport optimal, il existe $C \in \mathbb{R}$ tels que $\varphi_1 = \varphi_2 + C$ à l'intérieur du support de P. En particulier $\varphi_1 = \varphi_2 + C$ P-presque partout.*

Preuve. On a vu dans la preuve du théorème 4.2 que pour P-presque tout $x \in \mathbb{R}^d$, $\nabla\varphi_1 = \nabla\varphi_2$. Ainsi l'intérieur du support de P est un convexe sur lequel φ_1 et φ_2 sont finies (voir théorème 4.2) donc le lemme 4.5 apporte l'existence de $C \in \mathbb{R}$ tel que $\varphi_1 = \varphi_2 + C$ sur $\mathrm{int}(\mathrm{Supp}(P))$. Admettant le lemme 4.7 énoncé ci-dessous, $\lambda(\partial\mathrm{Supp}(P)) = 0$, donc on vérifie bien que $\varphi_1 = \varphi_2 + C$ P-presque partout. $\square$

Lemme 4.7. *Soit $C \subset \mathbb{R}^d$ un convexe d'intérieur non vide. Alors $\lambda(\partial C) = 0$.*

On peut énoncer deux théorèmes utiles pour la suite qui sont démontrés dans [BL19] (théorèmes 2.8 et 2.10). On rappelle pour justifier l'emploi de distances de Wasserstein qu'on a vu dans la preuve du théorème 2.18 que deux mesures admettant chacune un moment d'ordre p ont une distance de Wasserstein (l'une par rapport à l'autre) finie.

Théorème 4.8. *Soient P, Q, $(P_n)_n$, $(Q_n)_n$ des mesures de probabilités sur $\mathbb{R}^d$ de moments d'ordre deux finis. On suppose que Q est à support convexe, absolument continue par rapport à λ et que sa densité est strictement positive à l'intérieur de son support ainsi que $W_2(P_n, P) \to 0$ et $W_2(Q_n, Q) \to 0$. Soient ψ_n, potentiel de tranport optimal de Q_n à P_n et ψ de Q à P, alors il existe une suite $(a_n)_n \subset \mathbb{R}$ telle que pour tout x à l'intérieur de support de Q, $\psi_n(x) - a_n \to \psi(x)$, en particulier pour Q-presque tout $x \in \mathbb{R}^d$.*

C'est-à-dire que quand il y a convergence au sens de la distance de Wasserstein de $(P_n)_n$ vers P et de $(Q_n)_n$ vers Q implique la convergence des potentiels de transport (à constante près) vers le potentiel de transport de P à Q.

Théorème 4.9. *Soient P, Q, $(P_n)_n$ des mesures de probabilités sur $\mathbb{R}^d$ de moments d'ordre quatre finis. On suppose que Q est à support convexe, absolument continue par rapport à λ et que sa densité est strictement positive à l'intérieur de son support. Soient ψ_n, potentiel de tranport optimal de Q à P_n et ψ de Q à P, alors $\psi, \psi_n \in L^2(Q)$. De plus si $W_4(P_n, Q) \to 0$, $\varphi_n - a_n \to \varphi$ dans $L^2(Q)$, $(a_n)_n$ étant définie comme dans le théorème 4.8.*

Remarque 4.10. Il faut préciser en quoi ici on évoque une suite $(a_n)_n$ comme dans le théorème 4.8 bien que le cadre soit légèrement différent. En l'occurence la suite $(Q_n)_n$ est la suite $(Q)_n$ et vérifie bien $W_2(Q_n, Q) = W_2(Q, Q) = 0$. Ensuite l'hypothèse $W_4(P_n, Q) \to 0$ implique $W_2(P_n, Q) \to 0$, en effet par l'inégalité de Cauchy-Schwarz quelque soit $\pi \in \Pi(P_n, Q)$,

$$W_2(P_n, Q)^2 \leq \int_{\mathbb{R}^d} ||x-y||^2 d\pi(x,y) \leq \left(\int_{\mathbb{R}^d} ||x-y||^4 d\pi(x,y) \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^d} 1 d\pi(x,y) \right)^{\frac{1}{2}}$$

et ce quelque soit $\pi \in \Pi(P_n, Q)$ donc on peut (après passage à la racine carrée) passer à la borne inférieure à droite pour obtenir $W_2(P_n, Q) \leq W_4(P_n, Q)$. C'est pourquoi les hypothèses du théorème 4.8 sont bien vérifiées et l'existence d'une telle suite $(a_n)_n$ est justifée.

5 Théorème limite central pour la distance de Wasserstein quadratique entre une probabilité empirique et une probabilité cible

Définissons pour toute la suite $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé sur lequel seront définies les futures variables aléatoires. A partir de maintenant P et Q désignent des mesures de probabilité sur $\mathbb{R}^d$, $X_1, \ldots, X_n$ des variables aléatoires i.i.d (identiquement distribuées et indépendantes) de loi P et P_n désigne la mesure empirique des $X_1, \ldots, X_n$, c'est-à-dire que

$$P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} : \omega \in \Omega \mapsto \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}.$$

Définie ainsi, P_n est une mesure de probabilité aléatoire, c'est-à-dire une suite de variables aléatoires qui associent aux $\omega \in \Omega$, une mesure de probabilité $P_n(\omega)$. Dans ce cas, on montrera (dans la preuve de la proposition 5.1) que pour presque tout $\omega \in \Omega$, $P_n(\omega)$ converge en loi vers P ce qui fait comprendre que P_n se rapproche de P et donc qu'on peut chercher à décrire l'impact de ce phénomène sur la distance de Wasserstein entre P_n et P. Dès lors, les quantités d'intérêt telles que $W_2^2(P_n, P)$ et $W_2^2(P_n, Q)$ sont des variables aléatoires qui à chaque $\omega \in \Omega$ associent respectivement les quantités $W_2^2(P_n(\omega), P)$ et $W_2^2(P_n(\omega), Q)$.

5.1 Convergence presque sûre de la suite des distances de Wasserstein

Proposition 5.1. *Dans le cadre dans lequel nous nous situons,*

$$W_2^2(P_n, P) \xrightarrow[n\to+\infty]{p.s} 0 \text{ et } W_2^2(P_n, Q) \xrightarrow[n\to+\infty]{p.s} W_2^2(P, Q). \tag{41}$$

Preuve. Montrons d'abord la première convergence. Nous nous servirons de la proposition suivante, admise mais prouvée dans [GSZ18] (théorème 7.5), qui la couvre. On veut donc montrer que pour presque tout $\omega \in \Omega$, $W_2^2(P_n(\omega), P) \to 0$.

Proposition 5.2. *Soit μ et $(\mu_n)_n$ toutes des mesures de probabilités sur $\mathbb{R}^d$ de moments d'ordre $p \geq 1$ finis. Alors $W_p(\mu_n, \mu) \to 0$ si et seulement si μ_n converge faiblement vers μ et les moments d'ordre p de μ_n convergent vers ceux de μ.*

Première étape : presque sûre convergence des moments d'ordre deux
Tout d'abord,

$$\int_{\mathbb{R}^d} ||x||^2 dP_n(x) = \frac{1}{n}\sum_{i=1}^{n} ||X_i||^2 \xrightarrow[n\to+\infty]{\text{p.s}} \mathbb{E}\left[||X_1||^2\right] = \int_{\mathbb{R}^d} ||x||^2 dP(x), \tag{42}$$

convergence assurée par la loi forte des grands nombres car les $||X_i||^2$ sont des variables aléatoires i.i.d et intégrables.

Deuxième étape : presque sûre convergence en loi de la suite $(P_n)_n$ vers P

Remarque 5.3. La convergence que l'on cherche à prouver se traduit ainsi : pour presque tout $\omega \in \Omega$, $P_n(\omega) \xrightarrow{w} P$, c'est-à-dire que pour tout fonction $h : \mathbb{R}^d \to \mathbb{R}$ continue et bornée,

$$\int_{\mathbb{R}^d} h(x) dP_n(\omega)(x) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} h(x) dP(x). \tag{43}$$

Par la loi forte des grands nombres à nouveau, pour toute $h : \mathbb{R}^d \to \mathbb{R}$ continue et bornée,

$$\int_{\mathbb{R}^d} h(x)dP_n(\omega)(x) = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow[n\to+\infty]{\text{p.s}} \mathbb{E}[f(X_1)] = \int_{\mathbb{R}^d} h(x)dP(x)$$

c'est-à-dire que pour toute fonction test h, pour presque tout $\omega \in \Omega$, on vérifie (43). Ainsi, le négligeable implicite dans le "pour presque tout" est dépendant de la fonction test h. Tout l'enjeu est d'obtenir un négligeable qui ne depend pas de la fonction de test pour obtenir le résultat dans le bon ordre "pour presque tout ω, pour tout fonction h", l'ensemble des fonctions continues et bornées n'étant pas dénombrable. Chercher un ensemble dénombrable sur lequel (43) est vérifié permettra de résoudre ce problème.

Soit $\mathscr{C}_0(\mathbb{R}^d, \mathbb{R})$ l'ensemble des fonctions continues sur $\mathbb{R}^d$ à valeurs réelles tendant vers 0 à l'infini et $(f_k)_k$ une suite totale dans $\mathscr{C}_0(\mathbb{R}^d, \mathbb{R})$ c'est-à-dire que si $\Gamma := \text{Vect}(\{f_k; k \in \mathbb{N}\})$ alors $\overline{\Gamma} = \mathscr{C}_0(\mathbb{R}^d, \mathbb{R})$ (on munit ici $\mathscr{C}_0(\mathbb{R}^d, \mathbb{R})$ de la topologie de la norme infinie). $\{\exp(-a||x||^2 + b \cdot x); a \in \mathbb{Q}_+, b \in \mathbb{Q}^d\}$ en est une mais nous l'admettrons. Ainsi quelque soit $k \in \mathbb{N}$, par la loi forte des grands nombres

$$\int_{\mathbb{R}^d} f_k dP_n = \frac{1}{n}\sum_{i=1}^{n} f_k(X_i) \xrightarrow[n\to+\infty]{\text{p.s}} \int_{\mathbb{R}^d} f_k dP$$

donc pour tout $k \in \mathbb{N}$ il existe un ensemble négligeable N_k tel que pour tout $\omega \in \Omega \setminus N_k$,

$$\int_{\mathbb{R}^d} f_k dP_n(\omega) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f_k dP$$

et donc en posant $N' = \bigcup_{k\in\mathbb{N}} N_k$, $\mathbb{P}(N') = 0$ et pour tout $\omega \in \Omega \setminus N'$, pour tout $h_k \in \Gamma$ (par linéarité),

$$\int_{\mathbb{R}^d} h_k dP_n(\omega) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} h_k dP. \tag{44}$$

Soit $\omega \in \Omega \setminus N'$ fixé, $h \in \mathscr{C}_0(\mathbb{R}^d, \mathbb{R})$ et $(h_k)_k \subset \Gamma$ qui approche h en norme infinie. Alors pour $n \in \mathbb{N}$,

$$\begin{aligned}\left|\int_{\mathbb{R}^d} hdP_n(\omega) - \int_{\mathbb{R}^d} hdP\right| &\leq \left|\int_{\mathbb{R}^d}(h-h_k)dP_n(\omega) + \int_{\mathbb{R}^d}(h-h_k)dP\right| \\ &\quad + \left|\int_{\mathbb{R}^d} h_k dP_n(\omega) - \int_{\mathbb{R}^d} h_k dP\right| \leq 2||h-h_k||_\infty + \left|\int_{\mathbb{R}^d} h_k dP_n(\omega) - \int_{\mathbb{R}^d} h_k dP\right|\end{aligned}$$

donc par (44)

$$\limsup_{n\to+\infty}\left|\int_{\mathbb{R}^d} hdP_n(\omega) - \int_{\mathbb{R}^d} hdP\right| \leq 2||h-h_k||_\infty$$

or $||h-h_k||_\infty \to 0$ donc

$$\int_{\mathbb{R}^d} hdP_n(\omega) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} hdP.$$

et ce quelque soit $h \in \mathscr{C}_0(\mathbb{R}^d, \mathbb{R})$. Pour étendre ce résultat à toute fonction continue bornée, introduisons pour $x \in \mathbb{R}^d$ $g_k(x) = (x + (k+1))\mathbf{1}_{[-(k+1);-k]}(x) + \mathbf{1}_{[-k,k]}(x) + (-x+k+1)\mathbf{1}_{[k,k+1]}(x))$. Alors pour $h : \mathbb{R}^d \to \mathbb{R}$ continue et bornée,

$$\begin{aligned}\left|\int_{\mathbb{R}^d} hdP_n(\omega) - \int_{\mathbb{R}^d} hdP\right| &\leq \left|\int_{\mathbb{R}^d} hg_k dP_n(\omega) - \int_{\mathbb{R}^d} hg_k dP\right| + \left|\int_{\mathbb{R}^d} h(1-g_k)dP_n(\omega)\right| + \left|\int_{\mathbb{R}^d} h(1-g_k)dP\right| \\ &\leq \left|\int_{\mathbb{R}^d} hg_k dP_n(\omega) - \int_{\mathbb{R}^d} hg_k dP\right| + ||h||_\infty\left(\int_{\mathbb{R}^d}(1-g_k)dP_n(\omega) + \int_{\mathbb{R}^d}(1-g_k)dP\right) \\ &= \left|\int_{\mathbb{R}^d} hg_k dP_n(\omega) - \int_{\mathbb{R}^d} hg_k dP\right| + ||h||_\infty\left(2 - \int_{\mathbb{R}^d} g_k dP - \int_{\mathbb{R}^d} g_k dP_n(\omega)\right).\end{aligned}$$

Or $hg_k, g_k \in \mathscr{C}_0(\mathbb{R}^d, \mathbb{R})$ donc $\displaystyle\int_{\mathbb{R}^d} hg_k dP_n(\omega) \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} hg_k dP$ et $\displaystyle\int_{\mathbb{R}^d} g_k dP_n(\omega) \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} g_k dP$ donc

$$\limsup_{n\to+\infty} \left| \int_{\mathbb{R}^d} hdP_n(\omega) - \int_{\mathbb{R}^d} hdP \right| \leq 2||h||_\infty \left(1 - \int_{\mathbb{R}^d} g_k dP \right) \tag{45}$$

et par convergence monotone $\displaystyle\int_{\mathbb{R}^d} g_k dP \underset{k\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} 1dP = 1$ donc on peut déduire de (45) que (et ce quelque soit $h : \mathbb{R}^d \to \mathbb{R}$ continue et bornée)

$$\int_{\mathbb{R}^d} hdP_n(\omega) \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} hdP$$

donc $P_n(\omega)$ converge étroitement vers P pour $\omega \in \Omega \setminus N'$ quelconque. Ainsi, relativement au négligeable $N'' := N \cup N'$,

$$W_2(P_n, P) \xrightarrow[n\to+\infty]{\text{p.s}} 0$$

Egalement par inégalité triangulaire

$$|W_2(P_n, Q) - W_2(P, Q)| \leq W_2(P_n, P) \xrightarrow[n\to+\infty]{\text{p.s}} 0 \text{ ainsi } W_2(P_n, Q) \xrightarrow[n\to+\infty]{\text{p.s}} W_2(P, Q).$$

□

5.2 Théorème limite central

La convergence presque sûre étant établie on peut se poser les traditionnelles questions de la vitesse de la convergence et des fluctuations autour de la moyenne, c'est-à-dire d'un théorème limite central adapté à la situation. C'est le résultat principal de ce mémoire et il a été démontré aux alentours de 2017 par Eustasio del Barrio (Université de Valladolid, Espagne) et Jean-Michel Loubes (Université de Toulouse) puis été publié en 2019 dans la prestigieuse revue Annals of Probability, voir [BL19].

Théorème 5.4. *On suppose que P et Q sont toutes les deux à support convexe, absolument continues par rapport à λ et que leurs densités sont strictement positives à l'intérieur de leur support. On suppose également qu'elles sont de moment d'ordre $4 + \delta$ fini pour un certain $\delta > 0$. Soit φ_0 un potentiel de transport optimal de P à Q. Alors*

$$n\text{Var}\left(W_2^2(P_n, Q) - \int_{\mathbb{R}^d} (||x||^2 - 2\varphi_0(x)) dP_n(x) \right) \underset{n\to+\infty}{\longrightarrow} 0 \tag{46}$$

puis

$$\sqrt{n}\left(W_2^2(P_n, Q) - \mathbb{E}\left[W_2^2(P_n, Q) \right] \right) \xrightarrow[n\to+\infty]{w} \mathscr{N}\left(0, \sigma^2(P, Q) \right). \tag{47}$$

où

$$\sigma^2(P, Q) = \int_{\mathbb{R}^d} (||x||^2 - 2\varphi_0(x))^2 dP(x) - \left(\int_{\mathbb{R}^d} (||x||^2 - 2\varphi_0(x)) dP(x) \right)^2$$

Remarque 5.5. L'hypothèse particulière sur les moments d'ordre $4 + \delta$ finis est liée à des questions d'uniforme intégrabilité qui se soulèveront assez vite dans le déroulement de la preuve. En particulier elles s'expliquent par la proposition suivante ([Bre14], proposition 7.3.3 page 116)

Proposition 5.6. *Soit $p \geq 1$ $(Y_n)_n$ une suite de variables aléatoires dans $L^p(\mathbb{P})$, bornée dans $L^{p+\delta}(\mathbb{P})$ pour un certain $\delta > 0$. Alors la suite $(Y_n^2)_n$ est uniformément intégrable.*

5.2.1 Preuve du théorème limite central 5.4 - première étape

Enonçons l'inégalité principale dont on va se servir. Elle est admise mais on peut en trouver une preuve dans [BLM13] (théorème 3.1 page 54).

Théorème 5.7 (Inégalité d'Efron-Stein)**.** *Soient $Y_1, \dots, Y_n$ des variables aléatoires indépendantes et $Z = f(Y_1, \dots, Y_n)$ où $f : \mathbb{R}^d \to \mathbb{R}$ est une fonction telle que Z soit de carré intégrable. Si $Y_1', \dots, Y_n'$ sont des copies indépendantes de $Y_1, \dots, Y_n$ et pour tout i, $Z_i := f(Y_1, \dots, Y_{i-1}, Y_i', Y_{i+1}, \dots, Y_n)$, alors*

$$\operatorname{Var}(Z) \leq \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[(Z - Z_i)^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[(Z - Z_i)_+^2\right]$$

où $(Z - Z_i)_+ = \max\{Z - Z_i, 0\}$. En particulier si f est une fonction symétrique en les variables $x_1, \dots, x_n$, pour tout i, $\mathbb{E}\left[(Z - Z_i)_+^2\right] = \mathbb{E}\left[(Z - Z_1)_+^2\right]$ et donc

$$\operatorname{Var}(Z) \leq n\mathbb{E}\left[(Z - Z_1)_+^2\right].$$

On note $R_n := W_2^2(P_n, Q) - \int_{\mathbb{R}^d} (||x||^2 - 2\varphi_0(x)) dP_n(x)$. Par l'inégalité 5.7 d'Efron-Stein,

$$\operatorname{Var}(R_n) \leq n\mathbb{E}\left[(R_n - R_n')_+^2\right] \tag{48}$$

où R_n' est définie comme R_n mais pour P_n' qui est la mesure empirique de $X_1', X_2, \dots, X_n$ où X_1' est indépendante de $X_1, \dots, X_n$ et suit une loi P. Il suffit donc de montrer que $n^2\mathbb{E}\left[(R_n - R_n')_+^2\right] \to 0$, i.e que $(n(R_n - R_n')_+)_n$ converge dans L^2 vers 0. Pour cela il suffit de montrer que $n(R_n - R_n')_n \to 0$ p.s et que $\left(n^2(R_n - R_n')_+^2\right)_n$ est uniformément intégrable. En effet, la convergence p.s de $(n(R_n - R_n')_+)_n$ vers 0 implique la convergence en probabilités vers 0 de $\left(n^2(R_n - R_n')_+^2\right)_n$, puis on utilise la proposition suivante (admise mais démontrée dans [Bre14], théorème 7.3.1 page 117) avec $Y_n = n^2(R_n - R_n')_+^2$

Proposition 5.8. *Soit $(Y_n)_n$ une suite de variables aléatoires intégrables. Alors $(Y_n)_n$ converge dans L^1 si et seulement si elle est uniformément intégrable et converge en probabilités.*

Remarque 5.9. Rappelons que φ_0 est un potentiel de transport optimal de P à Q et ceci est équivalent au fait que

$$W_2^2(P, Q) = \int_{\mathbb{R}^d} ||x||^2 dP(x) + \int_{\mathbb{R}^d} ||y||^2 dQ(y) - 2\left(\int_{\mathbb{R}^d} \varphi_0(x) dP(x) + \int_{\mathbb{R}^d} \varphi_0^*(y) dQ(y)\right)$$

et dans ce cas

$$W_2^2(Q, P) = W_2^2(P, Q) = \int_{\mathbb{R}^d} ||y||^2 dQ(y) + \int_{\mathbb{R}^d} ||x||^2 dP(x) - 2\left(\int_{\mathbb{R}^d} \varphi_0^*(y) dQ(y) + \int_{\mathbb{R}^d} \varphi_0(x) dP(x)\right)$$

ce qui de la même manière est équivalent au fait que φ_0^* est un potentiel de transport optimal entre Q et P. De même si φ_n est un potentiel de transport entre P_n et Q, φ_n^* en est un entre Q et P_n. En l'occurence, remarquons qu'un potentiel de transport φ_n entre P_n et Q est plutôt une variable aléatoire $\omega \mapsto \varphi_n^{(\omega)}$ où chaque $\varphi_n^{(\omega)}$ est un potentiel de transport de $P_n(\omega)$ à Q.

Soit φ_n un potentiel de transport entre P_n et Q. Alors

$$\begin{aligned} W_2^2(P_n, Q) &= \int_{\mathbb{R}^d} ||x||^2 dP_n(x) + \int_{\mathbb{R}^d} ||y||^2 dQ(y) - 2\int_{\mathbb{R}^d} \varphi_n(x) dP_n(x) - 2\int_{\mathbb{R}^d} \varphi_n^*(y) dQ(y) \\ &= \int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_n(x)\right) dP_n(x) + \int_{\mathbb{R}^d} \left(||y||^2 - 2\varphi_n^*(y)\right) dQ(y). \end{aligned}$$

On note φ_n' un potentiel de transport entre P_n' et Q, alors de même

$$W_2^2(P_n', Q) = \int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_n'(x)\right) dP_n'(x) + \int_{\mathbb{R}^d} \left(||y||^2 - 2(\varphi_n')^*(y)\right) dQ(y) \tag{49}$$

Reprenant les notation introduites respectivement en (18) et (20) on note cette fois $J_{P_n',Q}$ et $\widetilde{\Phi}_{P_n',Q}$ et pour indiquer correctement les mesures en question. $(\varphi_n, \varphi_n^*) \in \widetilde{\Phi}_{P_n',Q}$: $\varphi_n \in L^1(P_n')$ car P_n' est une mesure finie, $\varphi_n^* \in L^1(Q)$ car on sait que $(\varphi_n, \varphi_n^*) \in \widetilde{\Phi}_{P_n,Q}$ et on sait que quelque soient $x, y \in \mathbb{R}^d$, $\varphi_n(x) + \varphi_n^*(y) \geq x \cdot y$. Ainsi

$$\int_{\mathbb{R}^d} \varphi_n' dP_n' + \int_{\mathbb{R}^d} (\varphi_n')^* dQ = \inf_{(\varphi,\psi)\in\widetilde{\Phi}_{P_n',Q}} J_{P_n',Q}(\varphi,\psi) \leq J_{P_n',Q}(\varphi_n, \varphi_n^*) = \int_{\mathbb{R}^d} \varphi_n dP_n' + \int_{\mathbb{R}^d} \varphi_n^* dQ$$

donc en combinant avec (49)

$$W_2^2(P_n', Q) \geq \int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_n(x)\right) dP_n'(x) + \int_{\mathbb{R}^d} \left(||y||^2 - 2\varphi_n^*(y)\right) dQ(y).$$

Cette inégalité permet d'obtenir

$$\begin{aligned}
R_n - R_n' = W_2^2(P_n, Q) - &\int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_0(x)\right) dP_n(x) - \left(W_2^2(P_n', Q) - \int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_0(x)\right) dP_n'(x)\right) \\
&\leq \int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_n(x)\right) dP_n(x) + \int_{\mathbb{R}^d} \left(||y||^2 - 2\varphi_n^*(y)\right) dQ(y) - \int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_0(x)\right) dP_n(x) \\
&- \int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_n(x)\right) dP_n'(x) - \int_{\mathbb{R}^d} \left(||y||^2 - 2\varphi_n^*(y)\right) dQ(y) + \int_{\mathbb{R}^d} \left(||x||^2 - 2\varphi_0(x)\right) dP_n'(x) \\
&= 2\int_{\mathbb{R}^d} \left(\varphi_0(x) - \varphi_n(x)\right) dP_n(x) - 2\int_{\mathbb{R}^d} \left(\varphi_0(x) - \varphi_n(x)\right) dP_n'(x).
\end{aligned}$$

Or $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ et $P_n' = \frac{1}{n}\delta_{X_1'} + \frac{1}{n}\sum_{i=2}^n \delta_{X_i}$ donc

$$R_n - R_n' \leq \frac{2}{n}\left(\left(\varphi_0(X_1) - \varphi_n(X_1)\right) - \left(\varphi_0(X_1') - \varphi_n(X_1')\right)\right) \tag{50}$$

Or on a vu dans le théorème 4.8 qu'il est éventuellement possible de faire converger les potentiels de transports. Par l'argument vu en proposition 5.1, $W_2(P_n, P) \to 0$ p.s relativement à un négligeable qu'on note N. On utilise alors le théorème 4.8, avec la mesure de probabilité P qui est à support convexe, absolument continue par rapport à λ, de densité strictement positive à l'intérieure de son support (on inverse en l'occurence P et Q dans le théorème 4.8). Les suites en questions sont $(P_n(\omega))_n$ et $(Q)_n$. L'application du théorème 4.8 s'écrit alors : pour tout $\omega \in \Omega \setminus N$, il existe une suite de potentiels de transport $(\varphi_n^{(\omega)})_n$ de $P_n(\omega)$ à Q tel que pour P-presque tout $x \in \mathbb{R}^d$, $\varphi_n^{(\omega)}(x) \to \varphi_0(x)$. On note donc $\varphi_n : \omega \in \Omega\setminus N \mapsto \varphi_n^{(\omega)}$ et associe 0 sur N. Ainsi pour P-presque tout x, $\varphi_n(x)$ converge presque sûrement vers φ_0. De (50) on déduit que

$$n(R_n - R_n')_+ \leq 2\left(\left(\varphi_0(X_1) - \varphi_n(X_1)\right) - \left(\varphi_0(X_1') - \varphi_n(X_1')\right)\right)_+ \tag{51}$$

or par ce qu'on vient d'établir, $\varphi_0(X_1) - \varphi_n(X_1) \to 0$ presque-sûrement étant donné que X_1 et de loi P, de même pour X_1'. Ainsi (51) octroie

$$n(R_n - R_n') \xrightarrow[n\to+\infty]{\text{p.s}} 0.$$

5.2.2 Preuve du théorème limite central 5.4 - uniforme intégrabilité

Il nous reste pour prouver (46) à montrer que $\left(n^2(R_n - R'_n)_+^2\right)_n$ est uniformément intégrable. Or $n(R_n - R'_n)$ vaut

$$n\left(W_2^2(P_n,Q) - \int_{\mathbb{R}^d}(||x||^2 - 2\varphi_0(x))dP_n(x) - W_2^2(P'_n,Q) + \int_{\mathbb{R}^d}(||x||^2 - 2\varphi_0(x))dP'_n(x)\right)$$

$$= n\left(W_2^2(P_n,Q) - W_2^2(P'_n,Q)\right) - n\left(\frac{1}{n}\sum_{i=1}^n\left(||X_i||^2 - 2\varphi_0(X_i)\right) - \frac{1}{n}\left(||X'_1||^2 - 2\varphi_0(X'_1)\right)\right.$$
$$\left. - \frac{1}{n}\sum_{i=2}^n\left(||X_i||^2 - 2\varphi_0(X_i)\right)\right)$$

donc

$$n(R_n - R'_n) = n\left(W_2^2(P_n,Q) - W_2^2(P'_n,Q)\right) - \left(\left(||X_1||^2 - 2\varphi_0(X_1)\right) - \left(||X'_1||^2 - 2\varphi_0(X'_1)\right)\right)$$

et

$$(n(R_n - R'_n)_+)^2 \leq \left(n\left(W_2^2(P_n,Q) - W_2^2(P'_n,Q)\right) - \left(\left(||X_1||^2 - 2\varphi_0(X_1)\right) - \left(||X'_1||^2 - 2\varphi_0(X'_1)\right)\right)\right)^2$$

Il suffit de prouver l'uniforme intégrabilité du membre de droite. Par le théorème 4.9 comme les mesures ont un moment d'ordre 4 fini, $\left(||X_1||^2 - 2\varphi_0(X_1)\right)$ et $\left(||X'_1||^2 - 2\varphi_0(X'_1)\right)$ sont de moment d'ordre 2 finis donc $\left(\left(||X_1||^2 - 2\varphi_0(X_1)\right) - \left(||X'_1||^2 - 2\varphi_0(X'_1)\right)\right)^2$ est intégrable et est ainsi uniformément intégrable. Prouvons un lemme d'uniforme intégrabilité

Lemme 5.10. *Soit $(A_n)_n$ une suite de variables aléatoires et (B_n) une suite de variables aléatoires telle que $(B_n^2)_n$ soit uniformément intégrables. Si $(A_n^2)_n$ est uniformément intégrable alors $((A_n + B_n)^2)_n$ est uniformément intégrable.*

Preuve. $((A_n + B_n)^2)_n$ est uniformément intégrable si et seulement si

$$\sup_{n\in\mathbb{N}}\mathbb{E}[(A_n+B_n)^2] < +\infty \text{ et } \forall\varepsilon>0, \exists\delta>0, \forall C\in\mathscr{F}, \mathbb{P}(C)<\delta \implies \sup_{n\in\mathbb{N}}\mathbb{E}\left[(A_n+B_n)^2\mathbf{1}_C\right] < \varepsilon.$$

Soit $\varepsilon > 0$, alors par hypothèse $(B_n^2)_n$ est uniformément intégrable donc il existe $\delta_1 > 0$ tel que pour $C \in \mathscr{F}$ avec $\mathbb{P}(C) \leq \delta_1$, pour tout $n \in \mathbb{N}$, $\mathbb{E}\left[B_n^2\mathbf{1}_C\right] < \varepsilon$. Par l'inégalité "$(a+b)^2 \leq 2(a^2+b^2)$",

$$\mathbb{E}\left[(A_n+B_n)^2\mathbf{1}_C\right] \leq 2\mathbb{E}\left[A_n^2 + B_n^2\mathbf{1}_C\right] = 2\mathbb{E}\left[A_n^2\mathbf{1}_C\right] + 2\mathbb{E}\left[B_n^2\mathbf{1}_C\right] \leq 2\mathbb{E}\left[A_n^2\mathbf{1}_C\right] + 2\varepsilon \tag{52}$$

Supposons que $(A_n^2)_n$ est uniformément intégrable, on peut trouver de même $\delta_2 > 0$ tel que pour $C \in \mathscr{F}$ avec $\mathbb{P}(C) \leq \delta_2$, pour tout $n \in \mathbb{N}$, $\mathbb{E}\left[A_n^2\mathbf{1}_C\right] \leq \varepsilon$. Ainsi (52) devient

$$\forall C \in \mathscr{F}, \mathbb{P}(C) \leq \min\{\delta_1,\delta_2\} \implies \forall n\in\mathbb{N}, \mathbb{E}\left[(A_n+B_n)^2\mathbf{1}_C\right] \leq 4\varepsilon. \tag{53}$$

D'autre part, toujours sous l'hypothèse que $(A_n^2)_n$ est uniformément intégrable, il existe $M_A > 0$ tel que quelque soit $n \in \mathbb{N}$, $\mathbb{E}\left[A_n^2\right] \leq M_A$ et de même il existe $M_B > 0$ qui vérifie la même chose pour $(B_n)_n$. Ainsi pour tout $n \in \mathbb{N}$,

$$\mathbb{E}\left[(A_n+B_n)^2\right] \leq 2\mathbb{E}\left[A_n^2\right] + 2\mathbb{E}\left[B_n^2\right] \leq 2M_A + 2M_B.$$

en combinant avec (53), $((A_n+B_n)^2)_n$ est uniformément intégrable. $\square$

Par le lemme 5.10 fraichement prouvé il suffit de montrer que $\left(n^2\left(W_2^2(P_n,Q)-W_2^2(P_n',Q)\right)^2\right)_n$ est uniformément intégrable. Or, $\nabla\varphi_n^*$ étant une application de transport optimale (hypothèses de régularité vérifiées) entre Q et P_n,

$$W_2^2(P_n,Q)=W_2^2(Q,P_n)=\int_{\mathbb{R}^d}||y-\nabla\varphi_n^*(y)||^2dQ(y). \tag{54}$$

Notons $C_i=\{y\in\mathbb{R}^d;\nabla\varphi_n^*(y)=X_i\}$. Alors comme $\nabla\varphi_n^*$ transporte Q sur P_n,

$$Q(C_i)=Q\left((\nabla\varphi_n^*)^{-1}(\{X_i\})\right)=P_n(\{X_i\})=\frac{1}{n}.$$

Les C_i sont presque-sûrement disjoints, en effet, si $y\in C_i\cap C_j$, alors $X_i=X_j$. Or, en notant f_{X_i} et f_{X_j} les densités respectives de X_i et X_j

$$\mathbb{P}(X_i=X_j)=\iint_{\mathbb{R}^d\times\mathbb{R}^d}\mathbf{1}_{x_i=x_j}f_{X_i}(x_i)f_{X_j}(x_j)dxdy=\int_{\mathbb{R}^d}f_{X_j}(x_j)\left(\int_{\{x_j\}}f_{X_i}(x_i)dx_i\right)dx_j=0$$

donc $\mathbb{P}(C_i\cap C_j\neq\emptyset)\leq\mathbb{P}(X_i=X_j)=0$, donc $\mathbb{P}(C_i\cap C_j\neq\emptyset)=0$ donc

$$Q\left(\bigcup_{i=1}^n C_i\right)=Q\left(\bigsqcup_{i=1}^n C_i\right)=\sum_{i=1}^n\frac{1}{n}=1$$

c'est pourquoi (54) peut devenir

$$W_2^2(P_n,Q)=\sum_{i=1}^n\int_{C_i}||y-\nabla\varphi_n^*(y)||^2dQ(y)=\sum_{i=1}^n\int_{C_i}||y-X_i||^2dQ(y).$$

On note de même $C_1'=\{y\in\mathbb{R}^d;\nabla(\varphi_n')^*(y)=X_1'\}$ et $C_i'=\{y\in\mathbb{R}^d;\nabla(\varphi_n')^*(y)=X_i\}$ pour $i\in\{2,\dots,n\}$, donc

$$W_2^2(P_n',Q)=\int_{C_1'}||y-X_1'||^2dQ(y)+\sum_{i=2}^n\int_{C_i'}||y-X_i||^2dQ(y). \tag{55}$$

Notons $G:\begin{cases}y\in C_i'\mapsto X_i, 1\leq i\leq n\\ 0 \text{ autrement}\end{cases}$ alors G transporte Q sur P_n, en effet pour $B\in\mathscr{B}(\mathbb{R}^d)$

$$\begin{aligned}G_{\#Q}(B)=\int_{\mathbb{R}^d}\mathbf{1}_B\circ G(y)dQ(y)=\sum_{i=1}^n\int_{C_i'}\mathbf{1}_B\circ G(y)dQ(y)=\sum_{i=1}^n\mathbf{1}_B(X_i)Q(C_i')\\ =\frac{1}{n}\sum_{i=1}^n\mathbf{1}_B(X_i)=P_n(B).\end{aligned}$$

Ainsi (une application de transport optimal existant selon le théorème de Brenier 4.2),

$$\begin{aligned}W_2^2(P_n,Q)=W_2^2(Q,P_n)=\inf_{T;\,T_{\#Q}=P_n}\left(\int_{\mathbb{R}^d}||y-T(y)||^2dQ(y)\right)\leq\int_{\mathbb{R}^d}||y-G(y)||^2dQ(y)\\ =\sum_{i=1}^n\int_{C_i'}||y-X_i||^2dQ(y). \quad (56)\end{aligned}$$

C'est pourquoi on peut combiner (55) et (56) pour obtenir

$$
\begin{aligned}
W_2^2(P_n,Q)-W_2^2(P_n',Q) &\leq \int_{C_1'}(||y-X_1||^2-||y-X_1'||^2)dQ(y)\\
&= \int_{C_1'}(||y-X_1||-||y-X_1'||)(||y-X_1||+||y-X_1'||)dQ(y)\\
&\leq \int_{C_1'}(||y-X_1-X_1'+X_1'||-||y-X_1'||)(||y||+||X_1||+||y||+||X_1'||)dQ(y)\\
&\leq ||X_1-X_1'||\int_{C_1'}(2||y||+||X_1||+||X_1'||)dQ(y) = ||X_1-X_1'||\left(\frac{1}{n}(||X_1||+||X_1'||)+2\int_{C_1'}||y||dQ(y)\right)
\end{aligned}
$$

avec $\frac{1}{n}=Q(C_1')$, qui fournit finalement

$$\left(n(W_2^2(P_n,Q)-W_2^2(P_n',Q))\right)^2 \leq ||X_1-X_1'||^2\left((||X_1||+||X_1'||)+2n\int_{C_1'}||y||dQ(y)\right)^2.$$

Or $||X_1-X_1'||^2(||X_1||+||X_1'||)^2$ est uniformément intégrable car les variables sont de moments d'ordre 4 finis. On se réfère à nouveau au lemme 5.10 pour affirmer qu'il suffit de prouver l'uniforme intégrabilité de la suite $\left(\left(n||X_1-X_1'||\int_{C_1'}||y||dQ(y)\right)^2\right)_n$. Nous voulons utiliser l'hypothèse sur les moments d'ordre $4+\delta$ finis pour prouver l'uniforme intégrabilité (Cf proposition 5.6). En utilisant l'inégalité de Cauchy-Schwarz,

$$\mathbb{E}\left[\left(n||X_1-X_1'||\int_{C_1'}||y||dQ(y)\right)^{2+\frac{\delta}{2}}\right] \leq \mathbb{E}\left[||X_1-X_1'||^{4+\delta}\right]^{\frac{1}{2}}\mathbb{E}\left[\left(n\int_{C_1'}||y||dQ(y)\right)^{4+\delta}\right]^{\frac{1}{2}} \tag{57}$$

Par l'inégalité de Hölder (avec $p=4+\delta$, $q=\frac{p}{p-1}=\frac{4+\delta}{3+\delta}$)

$$\left(\int_{C_1'}||y||dQ(y)\right)^{4+\delta} \leq \left(\left(\int_{C_1'}||y||^{4+\delta}dQ(y)\right)^{\frac{1}{4+\delta}}\left(\int_{C_1'}dQ(y)\right)^{\frac{3+\delta}{4+\delta}}\right)^{4+\delta} = \frac{1}{n^{3+\delta}}\int_{C_1'}||y||^{4+\delta}dQ(y).$$

la dernière quantité à droite étant finie par hypothèse sur Q. On en déduit

$$\mathbb{E}\left[\left(n\int_{C_1'}||y||dQ(y)\right)^{4+\delta}\right] \leq n\mathbb{E}\left[\int_{C_1'}||y||^{4+\delta}dQ(y)\right]. \tag{58}$$

Or pour i, les variables $X_1',X_2,\ldots,X_n$ étant identiquement distribuées et indépendantes, les variables aléatoires $\int_{C_1'}||y||^{4+\delta}dQ(y),\ldots,\int_{C_n'}||y||^{4+\delta}dQ(y)$ le sont également donc

$$\mathbb{E}\left[\int_{C_i'}||y||^{4+\delta}dQ(y)\right] = \mathbb{E}\left[\int_{C_1'}||y||^{4+\delta}dQ(y)\right]$$

et alors

$$n\mathbb{E}\left[\int_{C_1'}||y||^{4+\delta}dQ(y)\right] = \sum_{i=1}^{n}\mathbb{E}\left[\int_{C_i'}||y||^{4+\delta}dQ(y)\right] = \mathbb{E}\left[\int_{\mathbb{R}^d}||y||^{4+\delta}dQ(y)\right] = \int_{\mathbb{R}^d}||y||^{4+\delta}dQ(y). \tag{59}$$

Ainsi en combinant (57), (58) et (59), étant donné que P et Q sont de moments d'ordre $4+\delta$ finis, $\left(\left(n||X_1 - X_1'||\int_{C_1'}||y||dQ(y)\right)\right)_n$ est bornée dans $L^{2+\delta/2}(\mathbb{P})$ donc $\left(\left(n||X_1 - X_1'||\int_{C_1'}||y||dQ(y)\right)^2\right)_n$ est uniformément intégrable par la proposition 5.6. En conséquence $\left(\left(n(W_2^2(P_n,Q) - W_2^2(P_n',Q))\right)^2\right)_n$ est uniformément intégrable, puis $\left(n^2(R_n - R_n')_+^2\right)_n$ l'est également. L'inégalité d'Efron-Stein (déjà appliquée en (48)) fournit enfin (46) :

$$n\mathrm{Var}(R_n) \underset{n\to+\infty}{\longrightarrow} 0. \tag{60}$$

5.2.3 Preuve du théorème limite central 5.4 - conclusion

Notons $F_n = \int_{\mathbb{R}^d}(||x||^2 - 2\varphi_0(x))dP_n(x) = \frac{1}{n}\sum_{i=1}^{n}(||X_i||^2 - 2\varphi_0(X_i))$ ainsi $R_n = W_2^2(P_n,Q) - F_n$ et

$$\begin{aligned} n\mathrm{Var}(R_n) = n\mathbb{E}\left[(R_n - \mathbb{E}[R_n])^2\right] &= n\mathbb{E}\left[\left(W_2^2(P_n,Q) - F_n - \mathbb{E}\left[W_2^2(P_n,Q) - F_n\right]\right)^2\right] \\ &= \mathbb{E}\left[\left(\sqrt{n}\left(W_2^2(P_n,Q) - \mathbb{E}[W_2^2(P_n,Q)\right) - \sqrt{n}\left(F_n - \mathbb{E}[F_n]\right)\right)^2\right] \end{aligned} \tag{61}$$

et les variables aléaotoires étant identiquement distribuées,

$$\begin{aligned} \mathbb{E}[F_n] = \mathbb{E}\left[\int_{\mathbb{R}^d}(||x||^2 - 2\varphi_0(x))dP_n(x)\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(||X_i||^2 - 2\varphi_0(X_i))\right] \\ = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[||X_i||^2 - 2\varphi_0(X_i)\right] &= \mathbb{E}\left[||X_1||^2 - 2\varphi_0(X_1)\right] = \int_{\mathbb{R}^d}(||x||^2 - 2\varphi_0(x))dP(x) := F \end{aligned}$$

donc de (60) et (61) on déduit

$$\sqrt{n}\left(W_2^2(P_n,Q) - \mathbb{E}[W_2^2(P_n,Q)\right) - \sqrt{n}\left(F_n - F\right) \underset{n\to+\infty}{\overset{L^2}{\longrightarrow}} 0 \tag{62}$$

Or $(||X_i||^2 - 2\varphi_0(X_i))_i$ est une suite de variables aléatoires i.i.d et de carré intégrable (par hypothèse sur les $||X_i||^2$ et par le théorème 4.9) donc le théorème central limite réel classique apporte

$$\sqrt{n}(F_n - F) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(||X_i||^2 - 2\varphi_0(X_i)) - F\right) \underset{n\to+\infty}{\overset{\mathscr{L}}{\longrightarrow}} \mathscr{N}(0, \mathrm{Var}(||X_1||^2 - 2\varphi_0(X_1))). \tag{63}$$

avec

$$\begin{aligned} \mathrm{Var}(||X_1||^2 - 2\varphi_0(X_1)) = \mathbb{E}\left[(||X_1||^2 - 2\varphi_0(X_1))^2\right] - \mathbb{E}\left[||X_1||^2 - 2\varphi_0(X_1)\right]^2 = \\ \int_{\mathbb{R}^d}(||x||^2 - 2\varphi_0(x))^2dP(x) - \left(\int_{\mathbb{R}^d}(||x||^2 - 2\varphi_0(x))dP(x)\right)^2 := \sigma^2(P,Q). \end{aligned}$$

On énonce (non démontré ici) alors le lemme de Slutsky pour conclure

Lemme 5.11 (Slutsky)**.** *Soient (Y_n) et (Z_n) deux suites de variables aléatoires, Y une variable aléatoire et $c \in \mathbb{R}$ telles que $Y_n \xrightarrow[n\to+\infty]{\mathscr{L}} Y$ et $Z_n \xrightarrow[n\to+\infty]{\mathscr{L}} c$. Alors $(Y_n, Z_n) \xrightarrow[n\to+\infty]{\mathscr{L}} (Y, c)$, en particulier $Y_n + Z_n \xrightarrow[n\to+\infty]{\mathscr{L}} Y + c$.*

En l'occurence la convergence L^2 de (62) implique

$$\sqrt{n}\left(W_2^2(P_n, Q) - \mathbb{E}[W_2^2(P_n, Q)\right) - \sqrt{n}\,(F_n - F) \xrightarrow[n\to+\infty]{\mathscr{L}} 0$$

et en combinant par le lemme de Slutsky la convergence (63) alors

$$\begin{aligned}&\sqrt{n}\left(W_2^2(P_n, Q) - \mathbb{E}[W_2^2(P_n, Q)\right) = \\ &\quad\Big(\sqrt{n}\left(W_2^2(P_n, Q) - \mathbb{E}[W_2^2(P_n, Q)\right) - \sqrt{n}\,(F_n - F)\Big) + \Big(\sqrt{n}\,(F_n - F)\Big) \xrightarrow[n\to+\infty]{\mathscr{L}} \mathscr{N}(0, \sigma^2(P, Q)).\end{aligned}$$

ce qui est bien la convergence (47) voulue et le théorème 5.4 est prouvé.

5.3 Autres théorèmes limites centraux

On peut énoncer deux autres théorèmes limites centraux sur le coût de transport. Le premier concerne une comparaison entre deux mesures empiriques : $X_1, \dots, X_n$ désignent toujours des variables aléatoires i.i.d de loi P, P_n leur mesure empirique et on introduit $Y_1, \dots, Y_m$ des variables aléatoires i.i.d de loi Q ainsi que Q_m leur mesure empirique. On réfère sa preuve à l'article [BL19] (théorème 4.1)

Théorème 5.12. *On suppose que P et Q sont toutes les deux à support convexe, absolument continues par rapport à λ et que leurs densités sont strictement positives à l'intérieur de leur support. On suppose également qu'elles sont de moment d'ordre $4 + \delta$ fini pour un certain $\delta > 0$. Soit φ_0 un potentiel de transport optimal de P à Q et ψ_0 de Q à P. Pour faire tendre n et m vers l'infini en même temps on exige $\frac{n}{n+m} \xrightarrow[n,m\to+\infty]{} t \in]0,1[$. Alors*

$$\frac{nm}{n+m}\, Var\left(W_2^2(P_n, Q_m) - \int_{\mathbb{R}^d} (||x||^2 - 2\varphi_0(x))dP_n(x) - \int_{\mathbb{R}^d} (||y||^2 - 2\psi_0(y))dQ_m(y)\right) \xrightarrow[n,m\to+\infty]{} 0$$

puis

$$\sqrt{\frac{nm}{n+m}}\left(W_2^2(P_n, Q_m) - \mathbb{E}[W_2^2(P_n, Q_m)]\right) \xrightarrow[n,m\to+\infty]{w} \mathscr{N}\left(0, (1-t)\sigma^2(P, Q) + t\sigma^2(Q, P)\right)$$

où $\sigma^2(P, Q)$ est définie comme dans 5.4 et

$$\sigma^2(Q, P) = \int_{\mathbb{R}^d} (||y||^2 - 2\psi_0(y))^2 dQ(y) - \left(\int_{\mathbb{R}^d} (||y||^2 - 2\psi_0(y))dQ(y)\right)^2.$$

La preuve se faisant de manière plutôt similaire à celle du théorème 5.4. On peut aussi énoncer un autre théorème limite central, cette fois-ci la mesure P a un support fini et on peut réduire les hypothèses sur les moments de Q (on réfère sa preuve à [BL19], théorème 4.3)

Théorème 5.13. *On suppose que P est à support fini et que Q est absolument continue par rapport à λ de densité strictement positive à l'intérieur de sons support. On suppose égelement que Q est de moment d'ordre 4 fini. Alors*

$$\sqrt{n}\left(W_2^2(P_n, Q) - W_2^2(P, Q)\right) \xrightarrow[n\to+\infty]{w} \mathscr{N}\left(0, \sigma^2(P, Q)\right).$$

Le principe de le preuve est ici différent, en effet supposer que P est à support fini permet de décrire assez précisément le potentiel de transport et d'en déduire le théorème énoncé, ainsi il n'y a pas besoin de passer par des arguments plus abstraits d'uniforme intégrabilité et c'est pourquoi il n'est pas requis à Q d'être de moment d'ordre $4 + \delta$ fini mais seulement de moment d'ordre 4 fini.

Références

[Bil99] Patrick BILLINGSLEY. *Convergence of Probability Measures*. John Wiley & Sons,Inc., 1999.

[BL19] Eustasio del BARRIO et Jean-Michel LOUBES. "Central limit theorems for empirical transportation cost in general dimension". Dans : *Ann. Probab.* 47.2 (2019), p. 926-951.

[BLM13] Stéphane BOUCHERON, Gábor LUGOSI et Pascal MASSART. *Concentration inequalities : A nonasymptotic theory of independence*. Oxford university press, 2013.

[Bre14] Jean-Christophe BRETON. *Notes du cours de Fondements des Probabilités*. 2014. URL : `https://perso.univ-rennes1.fr/jean-christophe.breton/Fichiers/proba_base.pdf`.

[Che13] Patrick CHERIDITO. *Convex analysis (notes de cours)*. 2013. URL : `https://people.math.ethz.ch/~patrickc/CA2013.pdf`.

[DP40] Nelson DUNFORD et Billy James PETTIS. "Linear operations on summable functions". Dans : *Transactions of the American Mathematical Society* 47.3 (1940), p. 323-392.

[GSZ18] Nathael GOZLAN, Paul-Marie SAMSON et Pierre-André ZITT. *Notes de cours sur le transport optimal*. 2018. URL : `https://perso.math.u-pem.fr/samson.paul-marie/pdf/coursM2transport.pdf`.

[Jiř54] Miloslav JIŘINA. "Conditional probabilities on strictly separable σ-algebras". Dans : *Czechoslovak Math. J* 4.79 (1954), p. 372-380.

[Vil03] Cédric VILLANI. "Topics in Optimal Transportation". Dans : *American Mathematical Society* 58 (2003).