Appariement euclidien, théorème AKT et transport optimal

Nathan de Montgolfier et Léonie Kittel

24 juin 2022

Table des matières

1 Introduction

Notations

- Q désigne l'intervalle $]-\pi,\pi]$. Q^d est donc un cube de $\mathbb{R}^d$.
- On note $\mathbb{T}^d$ le tore muni de sa métrique euclidienne obtenu par quotient des faces opposées de Q^d.
- Si μ (resp. ν) est une mesure de probabilités sur X (resp. Y) :
  - $\Pi(\mu,\nu)$ désigne l'ensemble des mesures de probabilités sur $X \times Y$ dont les marginales sont μ et ν.
  - si $X = Y$ est un espace métrique muni de sa tribu borélienne et que μ et ν ont un premier moment fini, $\mathrm{W}_1(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_X d(x,y)\,\mathrm{d}\pi$ est la distance de Wasserstein entre μ et ν.
  - On note $\widetilde{\mathrm{W}_1}(\mu,\nu)$ la distance de Wasserstein sur les mesures de probabilités sur $\mathbb{T}^d$ (pour ne pas la confondre avec la distance de Wasserstein sur Q^d, car les mesures sur Q^d passent naturellement au quotient sur $\mathbb{T}^d$)
- Si f est une fonction lipschitzienne, alors , $\|f\|_{\mathrm{Lip}}$ désigne la constante de Lipschitz de f.
- Soit (a_n) et (b_n) deux suites réelles. Lorsqu'il existe $c > 0$ et $C > 0$ telle que $\forall n \in \mathbb{N}, ca_n \leq b_n \leq Ca_n$, on note $a_n = \Theta(b_n)$.
- Pour $m \in \mathbb{Z}^d$ On note $f_\mu(m) = \int_{Q^d} \mathrm{e}^{i\langle m,x\rangle}\,\mathrm{d}\mu(x)$ la transformée de Fourier de μ (on considère toujours la transformée de Fourier sur $\mathbb{Z}^d$).
- Pour $v \in \mathbb{R}^d$ (en particulier $\mathbb{Z}^d$), on note $|v|$ sa norme euclidienne canonique.

1.1 Présentation

L'objectif de ce mémoire est d'étudier le problème d'appariement euclidien et plus spécifiquement une preuve récente (2020) d'un théorème de 1984 dû à Ajtai, Komlós et Tusnády [2] appelé théorème AKT. Ce mémoire a été réalisé dans le cadre d'un projet de L3 à l'ENS Ulm au printemps 2022. Nous avons été encadrés par Djalil Chafaï que nous tenons à remercier pour sa disponibilité, ses conseils, son aide et sa motivation à nous convaincre qu'il est toujours possible de sauver une preuve. Ce mémoire nous a permis d'apprendre des mathématiques qui dépassent le cadre ce théorème : dualités issues de l'analyse fonctionnelle, transport optimal, fonctions c-concaves, approximation de Luzin, ...

Voici le problème qui nous intéresse. On considère deux ensembles de n points de l'espace euclidien $\mathbb{R}^d$, $X = \{x_1, \ldots, x_n\}$ et $Y = \{y_1, \ldots, y_n\}$. Le problème d'appariement euclidien optimal consite à relier chaque point de X à un unique point de Y en cherchant à minimiser la somme totale des longueurs des traits tracés. Ce problème apparaît naturellement dans des situations concrètes : on peut imaginer un problème d'optimisation de livraisons par exemple. D'un point de vue algorithmique, c'est un cas particulier du problème de couplage minimal dans un graphe bipartite pondéré et des algorithmes pour le résoudre sont connus (utilisant par exemple des techniques de flots ou encore de programmation linéaire). Dans ce mémoire, nous laissons de côté les aspects algorithmiques et nous intéressons au côut de la solution optimale lorsque les points sont tirés aléatoirement. Ainsi, le théorème AKT quantifie l'asymptotique (quand le nombre de points est grand) de ce coût lorsque les points sont tirés selon une loi uniforme sur $[0,1]^d$. Cette question a priori probabiliste intéresse aussi des physiciens-statisticens (l'article de Caracciolo et Parisi [6] s'intéresse par exemple à la constante optimale dans le théorème). Nous présentons une preuve récente publiée par Bobkov et Ledoux dans [4] qui suit les idées de Ambrosio, Stra et Trevisan dans [3]. Cette preuve est basée sur des techniques d'analyse de Fourier et d'approximation de fonctions. Pour la comprendre, nous devons d'abord expliquer le lien du problème d'appariement euclidien avec le problème de transport optimal, c'est l'objet de la section 2. La section 3 présente la preuve du théorème AKT. Les sections 4 et 5 présentent deux outils d'analyse fonctionnelle qui nous ont été utiles dans la preuve : le théorème de Kantorovitch-Rubinstein et un lemme d'approximation par des fonctions lipschitziennes.

1.2 Enoncé du théorème AKT

Théorème 1.2.1 (AKT)**.** *Soit $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ des variables aléatoires indépendantes suivant une loi uniforme sur $[0,1]^d$. Alors,*

$$\mathbb{E}\left(\inf_{\sigma \in S_n} \frac{1}{n} \sum_{i=1}^{n} \left|X_i - Y_{\sigma(i)}\right|\right) = \begin{cases} \Theta\left(\frac{1}{\sqrt{n}}\right) & si\ d = 1 \\ \Theta\left(\sqrt{\frac{\log(n)}{n}}\right) & si\ d = 2 \\ \Theta\left(\frac{1}{n^{\frac{1}{d}}}\right) & si\ d \geq 3 \end{cases} \tag{1.2.1}$$

L'ordre de grandeur obtenu pour $d \geq 3$ est peu surprenant : tout se passe comme si chaque X_i était relié à son plus proche voisin dans $\{Y_1, \ldots, Y_n\}$ qui est typiquement dans un cube de volume $\frac{1}{n}$ et donc de diamètre $\Theta(n^{-\frac{1}{d}})$. En dimension 1, la solution explicite du problème est très simple et la borne supérieure s'obtient assez facilement. En dimension 2, un phénomène particulier se produit, c'est le cas le plus délicat.

Le théorème peut s'étendre à des familles de lois plus générales. Ainsi, on peut montrer que l'ordre de grandeur du théorème AKT est une borne supérieure pour toutes les lois de $(X_1, \dots, X_n, Y_1, \dots, Y_n)$ invariantes par permutations et à support dans $[0,1]^d$ (voir [4, §5]).

2 Le problème de transport optimal

2.1 Le problème de transport optimal

Le problème du transport optimal est le suivant. On dispose d'un stock de ressources toutes identiques situées en différents points de l'espace que l'on souhaite acheminer en différents points, on peut imaginer un tas de sable dont la hauteur varie dans l'espace que l'on souhaite déplacer vers un trou de même volume que le tas dont la profondeur varie dans l'espace. Le transport de la ressource a un coût, variable selon le point de départ et le point d'arrivée, et linéaire en la quantité de ressources acheminée. Le transport optimal s'intéresse à la minimisation du coût du transport global.

Plus formellement, fixons X un espace mesurable. On dispose de deux mesures de probabilités : une mesure de départ μ et une mesure d'arrivée ν (représentant donc respectivement le sable et le trou) ainsi que d'une fonction mesurable $c : X \times X \to \mathbb{R}_+ \cup \infty$, dite de coût. Un plan de transport au sens de Monge est donc une fonction mesurable $f : X \to X$ qui à chaque point x de X associe un point $f(x)$ vers lequel les ressources en x sont transportées. Le plan de transport doit vérifier la contrainte suivante : la mesure image de μ par f (notée $f(\mu)$) doit être ν, le sable doit combler exactement le trou. Le coût de transport de f est alors :

$$\int_X c(x, f(x))\, \mathrm{d}\mu(x)\,.$$

On peut alors poser deux problèmes de transport optimal différents : le problème de Monge et le problème de Kantorovitch. On appelle problème du transport optimal au sens de Monge le problème de minimisation de $\int_X c(x, f(x))\, \mathrm{d}\mu(x)$ sur l'ensemble des fonctions mesurables sur X.

Un plan de transport f donne de manière naturelle une mesure de probabilité π_f

sur $X \times X$ dont les marginales sont μ et ν : c'est la mesure définie par

$$\int\limits_{X\times X} g\,\mathrm{d}\pi_f = \int\limits_{X\times X} g(x, f(x))\,\mathrm{d}\mu(x)$$

pour toute fonction mesurable positive g.

Un plan de transport au sens de Kantorovitch est précisément la donnée d'une mesure de probabilité sur $X \times X$ dont les marginales sont μ et ν. On note $\Pi(\mu, \nu)$, l'ensemble des telles mesures de probabilité. La minimisation du coût de transport $\int_X c\,\mathrm{d}\pi$ pour π parcourant $\Pi(\mu, \nu)$ est le problème de Kantorovitch. Notons que c'est un problème de minimisation linéaire.

En terme de tas de sable, le problème de Kantorovitch autorise à déplacer des grains de sable initialement au même endroit vers des points d'arrivée différents alors que le problème de Monge contraint le point de livraison de chaque grain à une fonction du point d'arrivée.

Le coût minimal du problème de Kantorovitch est donc toujours majoré par celui du problème de Monge ; dans certains cas, comme celui du problème d'appariement euclidien (cf section 2.2), ces deux problèmes ont les mêmes solutions.

Le problème de Kantorovitch a l'avantage technique de permettre l'utilisation d'outils d'analyse fonctionnelle (dualité, optimisation convexe). Par exemple, si X est compact et c continue, le théorème de Banach-Alaoglu-Bourbaki permet de voir $\Pi(\mu, \nu)$ comme une partie faiblement-$*$ compacte du dual des fonctions continues munies de la topologie de la convergence uniforme, ce qui garantit l'existence d'un plan de transport optimal.

2.2 Lien avec le problème d'appariement

Le problème d'appariement s'exprime naturellement comme un problème de Monge. La mesure de départ est la mesure empirique de $\{x_1, \dots x_n\}$: $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Celle d'arrivée est la mesure empirique de $\{y_1, \dots, y_n\}$. La fonction de coût est simplement la distance sur $\mathbb{R}^d$. Une solution de ce problème est aussi une solution du problème de Kantorovitch associé. En effet, ici un plan de transport au sens de Kantorovitch peut être représenté par une matrice P dont le coefficient $P_{i,j}$ représente la masse à déplacer de x_i vers y_j. Les matrices P correspondant à des plans de transport au sens de Kantorovitch sont donc exactement les matrices telles que nP est bistochatique. Notons $\mathcal{P}$ cet ensemble et remarquons que $\mathcal{P}$ est convexe.

Nous devons donc minimiser $\displaystyle\sum_{1\leq i\leq n, 1\leq j\leq n} P_{i,j}\, d(x_i, x_j)$ pour P parcourant le convexe $\mathcal{P}$. Cette quantité est linéaire en P et atteint donc son minimum sur les points extrêmaux de $\mathcal{P}$. Par un théorème de Birkoff et Von Neumann sur les matrices bistochastiques, ces points sont précisément les matrices de permutation (avec un facteur $\frac{1}{n}$), c'est-à-dire les plans de transport qui déplacent chaque point x_i vers un unique point y_j et donc les solutions du problème de Monge.

Ainsi,

$$\inf_{\sigma \in S_n} \frac{1}{n} \sum_{i=1}^{n} \left| X_i - Y_{\sigma(i)} \right| = \inf_{\pi \in \Pi(\mu_n, \nu_n)} \int_{Q^d} d(x,y)\, \mathrm{d}\pi(x,y) := \mathrm{W}_1\left(\mu_n, \nu_n\right) \tag{2.2.1}$$

Si X est un espace métrique, la quantité $\mathrm{W}_1\left(\mu, \nu\right) = \inf_{\pi \in \Pi(\mu, \nu)} \int_X d(x,y)\, \mathrm{d}\pi(x,y)$ définit une distance sur l'espace des mesures de premier moment fini sur X. C'est la distance de Wasserstein.

3 Le théorème AKT

3.1 Approche

Cette section résume l'approche que nous suivons pour démontrer le théorème AKT.

Le théorème suivant (dû à Kantorovitch) fournit une stratégie pour démontrer le théorème AKT :

Théorème 3.1.1 (Théorème de Kantorovitch-Rubinstein)**.**

$$\mathrm{W}_1\left(\mu, \nu\right) = \sup_{\|f\|_{Lip} \leq 1} \int_X f\, \mathrm{d}(\mu - \nu) = \sup_{\|f\|_{Lip} \leq 1} \mathbb{E}\left(f(X)\right) - \mathbb{E}\left(f(Y)\right) \tag{3.1.1}$$

où $\|f\|_{Lip}$ désigne la constante de Lipschitz de f et où X et Y sont des variables aléatoires de lois respectives μ et ν.

Pour obtenir la borne supérieure, il suffit donc de contrôler $\mathbb{E}\left(f(X) - f(Y)\right)$ pour toutes les fonctions f 1-lipschitziennes. La preuve que nous suivons utilise des outils d'analyse de Fourier. Cette dernière se prête mieux à l'étude de fonctions très régulières mais ce n'est pas un problème ici puisque les fonctions $\mathcal{C}^\infty$ sont denses dans l'ensemble des fonctions 1-lipschitziennes[1]. En effet, le caractère lipschitzien d'une

1. Avec les notations de la section 3.2, si f est 1-Lipschitzienne alors, $\sigma_t * f$ est toujours 1-lipschitzienne. De plus si f est périodique $\sigma_t * f$ l'est aussi, et l'argument fonctionne donc aussi sur le tore.

fonction donne un contrôle explicite des coefficients de sa transformée de Fourier. Ce contrôle, combiné à un argument de régularisation des mesures, permet d'obtenir la borne supérieure du théorème AKT.

Pour obtenir une borne inférieure, il suffit a priori de construire une fonction lipschitzienne (aléatoire) qui se rapproche de la borne souhaitée. Si c'est élémentaire pour $d \geq 3$, une telle construction est délicate pour $d = 2$. Ajtai, Komlós et Tusnády, dans la preuve originale du théorème AKT [2], construisent une fonction naïve, puis la modifient pour qu'elle ait la bonne constante de Lipschitz en utilisant des arguments probabilistes reposant sur des ponts Browniens et le théorème d'approximation de Komlós-Major-Tusnády (théorème KMT).

L'approche de Bobkov et Ledoux est la suivante : après régularisation des mesures, pour une fonction lipschitzienne u, $\int_{Q^d} u \, \mathrm{d}(\mu - \nu) = \int_{Q^d} u(x) g(x) \, \mathrm{d}x$ où g est la densité de $\nu - \mu$. On réalise alors que $g = \Delta h$ pour une certaine fonction h et une intégration par partie donne donc $\int_{Q^d} u \, \mathrm{d}(\mu - \nu) = \int_{Q^d} \langle \nabla u(x), \nabla h(x) \rangle \, \mathrm{d}x$. Ainsi, pour maximiser cette quantité, il faut approcher h par une fonction 1-lipschitzienne. Ceci peut être fait grâce à un résultat quantitatif d'approximation des fonctions par des fonctions lipschitziennes.

3.2 Régularisation des mesures

Nous décrivons ici le procédé de régularisation des mesures que l'on utilisera dans la suite. Sur $\mathbb{R}^d$, convoler une mesure μ avec une gaussienne dont on peut faire tendre la variance vers 0 est une manière standard de régulariser la mesure μ. Pour $t > 0$ on considère η_t une gaussienne sur $\mathbb{R}$ de variance $2t$ et on note $\sigma_t = \eta_t^{\otimes d}$ la mesure produit itérée d fois de cette gaussienne. Ainsi, sa densité est radiale et s'écrit $f(r) = \frac{1}{(4\pi t)^{\frac{d}{2}}} \mathrm{e}^{-\frac{r^2}{4t}}$. On pourrait simplement convoler les mesures avec cette gaussienne, mais elles ne seraient alors plus à support dans Q^d.

Pour contourner ce problème, on pousse en avant la gaussienne sur Q vu comme représentant le tore $\mathbb{R}/(\pi + 2\pi\mathbb{Z})$. Pour ce faire, on considère la mesure $\tilde{\eta}_t$, mesure image par la projection P de $\mathbb{R}$ sur Q définie par $P(x) = x - 2k(x)\pi$ où $k(x)$ est l'entier tel que $(2k(x) - 1)\pi \leq x \leq (2k(x) + 1)\pi$.

La transformée de Fourier (à coefficients dans $\mathbb{Z}$) de $\tilde{\eta}_t$ sur Q est égale à la transformée de Fourier de η sur $\mathbb{R}$: en effet, pour $m \in \mathbb{Z}$,

$$f_{\tilde{\eta}_t}(m) = \int_{\mathbb{R}} \mathrm{e}^{imP(x)} \sigma_t(x) \, \mathrm{d}x = \int_{\mathbb{R}} \mathrm{e}^{imx} \sigma_t(x) \, \mathrm{d}x$$

par $2i\pi$ périodicité de l'exponentielle.

On note alors $\gamma_t = \tilde{\eta}^{\otimes d}$ la mesure produit sur Q^d, sa densité est

$$\rho_t(x) = \frac{1}{(2\pi)^d} \sum_{m \in \mathbb{Z}^d} \mathrm{e}^{i\langle m,x\rangle - m^2 t}.$$

Si μ est une mesure sur Q^d, on note $\mu_t = \mu * \gamma_t$ la convolée de μ et γ_t. Cette mesure est absolument continue par rapport à la mesure de Lebesgue, à densité $\mathcal{C}^\infty$. Remarquons que $f_{\mu_t}(m) = e^{-mt^2} f_\mu(m)$. En particulier, μ_t converge en loi vers μ lorsque $t \to 0$.

3.3 Borne supérieure

Dans les deux sections suivantes, $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ sont des variables aléatoires sur Q^d le cube, de loi uniforme $\mu = \nu = \mathbb{P}$, $\mu_n = \sum_{i=1}^n \delta_{X_i}$ et $\nu_n = \sum_{i=1}^n \delta_{Y_i}$ sont les mesures empiriques de $X_1, \ldots, X_n$ (resp. $Y_1, \ldots, Y_n$).

La démonstration de [4] repose sur le lemme suivant d'analyse de Fourier :

Lemme 3.3.1. *Soit $u \in \mathcal{C}^\infty$ une fonction 1-lipschitzienne sur $\mathbb{T}^d$. On écrit $u(x) = \sum_{m \in \mathbb{Z}^d} a_m \mathrm{e}^{i\langle m,x\rangle}$ sa décomposition en série de Fourier. Alors, $\sum_{m \in \mathbb{Z}^d} |m a_m|^2 \leq 1$*

Démonstration. En sommant l'identité de Parseval écrite sur chaque dérivée partielle, on obtient $\sum_{m \in \mathbb{Z}^d} |m a_m|^2 = \int_{Q^d} |\nabla u|^2$ et le caractère lispchitzien garantit $|\nabla u| \leq 1$. $\square$

On peut alors écrire :

$$\int_{Q^d} u \,\mathrm{d}(\mu - \nu) = \sum_{m \in \mathbb{Z}^d} a_m (f_\mu(m) - f_\nu(m))$$

(le terme en $m = 0$ est nul car $f_\mu(0) = f_\nu(0) = 1$)

Ainsi par l'inégalité de Cauchy-Schwartz,

$$\begin{aligned}
\left| \int_{Q^d} u \,\mathrm{d}(\mu - \nu) \right|^2 &\leq \left(\sum_{m \neq 0} |m a_m|^2 \right) \left(\sum_{m \neq 0} \frac{|f_\mu(m) - f_\nu(m)|^2}{|m|^2} \right) \\
&\leq \sum_{m \neq 0} \frac{|f_\mu(m) - f_\nu(m)|^2}{|m|^2} && \text{par le lemme 3.3.1}
\end{aligned}$$

La borne ne dépendant plus de u, on obtient la borne suivante (pas nécessairement finie) :

$$\widetilde{\mathrm{W}}_1\left(\mu,\nu\right)^2 \leq \sum_{m\neq 0} \frac{\left|f_\mu(m) - f_\nu(m)\right|^2}{\left|m\right|^2} \tag{3.3.2}$$

Cette borne s'applique également à $\mathrm{W}_1\left(\mu,\nu\right)$ dès lors que le support de μ et ν est inclus dans l'hypercube $]-\frac{\pi}{2},\frac{\pi}{2}]$ (sur cet hypercube plus petit, la distance sur le tore coïncide avec celle sur $\mathbb{R}^d$ car les géodésiques sur le tore entre les points de cet hypercube ne traversent jamais le bord de Q^d).
Pour $d = 1$, cette inégalité appliquée directement aux mesures empiriques suffit directement à prouver le théorème.
Pour $d \geq 2$ la série de l'équation (3.3.2) ne converge pas. En effet, pour $m \neq 0$, $\mathbb{E}\left(\left|f_{\mu_n}(m) - f_{\nu_n}(m))\right|^2\right) = \frac{2}{n}$

La régularisation des mesures permet ici de forcer la convergence de la somme. En effet pour $t > 0$,

$$\begin{aligned}
\widetilde{\mathrm{W}}_1\left(\mu_{n,t},\nu_{n,t}\right)^2 \leq \sum_{m\neq 0} \frac{\left|f_{\mu_{n,t}}(m) - f_{\nu_{n,t}}(m)\right|^2}{\left|m\right|^2} &= \sum_{m\neq 0} \frac{\mathrm{e}^{-2|m|^2 t}\left|f_{\mu_n}(m) - f_{\nu_n}(m)\right|^2}{\left|m\right|^2} \\
&\leq \frac{2}{n}\sum_{m\neq 0} \frac{\mathrm{e}^{-2|m|^2 t}}{\left|m\right|^2}
\end{aligned} \tag{3.3.3}$$

Lemme 3.3.4. *L'inégalité suivante permet de contrôler la distance entre une mesure de probabilités μ et sa régularisée :*

$$\widetilde{\mathrm{W}}_1\left(\mu_t,\mu\right) \leq \sqrt{2dt} \tag{3.3.4}$$

Démonstration. Elle résulte du calcul suivant :

$$\begin{aligned}
\widetilde{\mathrm{W}}_1\left(\mu_t,\mu\right) &= \sup_{\|u\|_{\mathrm{Lip}}\leq 1} \int_{Q^d}\int_{Q^d} \left(u(x-y) - u(x)\right)\rho_t(y)\,\mathrm{d}y\,\mathrm{d}\mu(x) \\
&\leq \int_{Q^d} |y|\rho_t(y)\,\mathrm{d}y = \mathbb{E}\left(|\gamma_t|\right) \\
&\leq \sqrt{\mathbb{E}\left(|\gamma_t|^2\right)} = \sqrt{2dt}
\end{aligned}$$

□

Ainsi, en combinant l'équation (3.3.3) et le lemme 3.3.4, on obtient le contrôle en espérance suivant :

Proposition 3.3.5.

$$\mathbb{E}\left(\widetilde{\mathrm{W}}_1\left(\mu_n,\nu_n\right)\right) \leq \sqrt{\frac{2}{n}\sum_{m\neq 0}\frac{\mathrm{e}^{-2|m|^2 t}}{|m|^2}} + 2\sqrt{2dt} \tag{3.3.5}$$

Par comparaison série intégrale, la somme dans le membre de droite est du même ordre de grandeur que l'intégrale suivante (lorsque $t \to 0$) :

$$\begin{aligned}\int\limits_{|x|\geq 1} \frac{1}{x^2}e^{-2|m|^2 t} &= V\int\limits_1^\infty \frac{1}{r^2}e^{-2tr|m|^2}r^{d-1}\,\mathrm{d}r \qquad \text{où } V \text{ est la mesure de } \mathbb{S}^{d-1} \\ &= \frac{1}{t^{\frac{d-2}{2}}}\int\limits_{\sqrt{t}}^\infty u^{d-3}\mathrm{e}^{-2u^2}\,\mathrm{d}u \qquad (3.3.6)\end{aligned}$$

Pour $d \geq 3$, l'intégrale est absolument convergente sur $\mathbb{R}_+$ et converge donc vers une constante lorsque $t \to 0$. Pour $d = 2$,

$$\int\limits_{\sqrt{t}}^\infty u^{-1}\mathrm{e}^{-2u^2}\,\mathrm{d}u \sim \int\limits_{\sqrt{t}}^1 u^{-1}\,\mathrm{d}u = \ln\!\left(\sqrt{t}\right)$$

Enfin, si $d = 1$,

$$\int\limits_{\sqrt{t}}^\infty u^{-2}\mathrm{e}^{-2u^2}\,\mathrm{d}u \sim \int\limits_{\sqrt{t}}^1 u^{-2}\,\mathrm{d}u = \frac{1}{\sqrt{t}} - 1$$

Ainsi quand $t \to 0$,

$$\int\limits_{|x|\geq 1} \frac{1}{x^2}e^{-2|m|^2 t} = \begin{cases}\Theta(1) & \text{si } d = 1\\ \Theta(\ln(t)) & \text{si } d = 2\\ \Theta(t^{-\frac{d}{2}+1}) & \text{si } d \geq 3\end{cases} \tag{3.3.7}$$

En optimisant le paramètre t en fonction n, pour que les deux termes de l'inégalité sur $\mathbb{E}\left(\widetilde{\mathrm{W}}_1\left(\mu_n,\nu_n\right)\right)$ soient du même ordre, $t(n) = \frac{1}{n}$ pour $d \geq 3$ et $t(n) = n^{-\frac{1}{d}}$, on obtient :

$$\mathbb{E}\left(\widetilde{\mathrm{W}_1}\left(\mu_n,\nu_n\right)\right) \leq \begin{cases} \Theta(\frac{1}{\sqrt{n}}) & \text{si } d=1 \\ \Theta(\sqrt{\frac{\ln(n)}{n}}) & \text{si } d=2 \\ \Theta(n^{-\frac{1}{d}}) & \text{si } d\geq 3 \end{cases} \tag{3.3.8}$$

3.4 Borne inférieure

3.4.1 En dimension supérieure à 3

Le cas $d \geq 3$ demande assez peu de travail : relier chaque point à son plus proche voisin donne le bon ordre de grandeur. Cette heuristique échoue pour $d = 1$ ou $d = 2$.

Par l'inégalité de Jensen, $\mathbb{E}\left(\mathrm{W}_1\left(\mu_n,\nu_n\right)\right) \geq \mathbb{E}\left(\mathrm{W}_1\left(\mu_n,\mathbb{P}\right)\right)$ On pose alors la fonction 1-lipschitzienne suivante : $f(x) = d(x, \{X_1,\dots,X_n\})$.
Ainsi,

$$\mathrm{W}_1\left(\mu_n,\mathbb{P}\right) \geq \left|\int f\ \mathrm{d}(\mu_n - \mathbb{P})\right| = \int d(x,\{X_1,\dots,X_n\})\ \mathrm{d}\mathbb{P} \tag{3.4.1}$$

Notons $\mathcal{C}$, l'ensemble des petits cubes obtenus en découpant $[0,1]^d$ en cubes de coté $\frac{1}{n^{\frac{1}{d}}}$.
Si $C \in \mathcal{C}$ est un cube, on note $\mathcal{V}(C)$ l'ensemble des $2d+1$ cubes voisins (C compris), alors

$$\mathbb{P}\left(\forall k \in \{1,2,\dots,n\}\, X_k \notin \mathcal{V}(C)\right) = \left(1 - \frac{2d+1}{n}\right)^n \to_{n\to\infty} \mathrm{e}^{-2d+1}. \tag{3.4.2}$$

En particulier, il existe une constante a ne dépendant que de d telle que $\forall n \in \mathbb{N}, \mathbb{P}\left(\forall k \in \{1,2,\dots,n\},\ X_k \notin C\right) > a$.

Ainsi,

$$\begin{aligned} \mathbb{E}\left(\mathrm{W}_1\left(\mu_n,\mathbb{P}\right)\right) &\geq \sum_{C\in\mathcal{C}} \int_C d(x,\{X_1,\dots,X_n\})\ \mathrm{d}\mathbb{P}(x) \\ &\geq \sum_{C\in\mathcal{C}} \mathbb{E}\left(d(C,\{X_1,\dots,X_n\}\right)\mathbb{P}(C) \\ &\geq \sum_{C\in\mathcal{C}} \frac{1}{n^{\frac{1}{d}}}\mathbb{P}\left(\forall k\in\{1,2,\dots,n\}\, X_k\notin\mathcal{V}(C)\right)\ \mathbb{P}(C) \\ &\geq \frac{a}{n^{\frac{1}{d}}}. \end{aligned} \tag{3.4.3}$$

3.4.2 En dimension 2

On commence par remarquer la minoration :

$$\widetilde{\mathrm{W}}_1\left(\mu,\nu\right) \geq \widetilde{\mathrm{W}}_1\left(\mu_t,\nu_t\right)$$

En effet, pour tout u,

$$\begin{aligned}
\int u\,\mathrm{d}(\mu_t-\nu_t) &= \int\limits_y \int\limits_x u(x-y)\,\mathrm{d}(\mu-\nu)\,(x)p_t(y)\,\mathrm{d}y \\
&= \int\limits_y \int v_y\,\mathrm{d}(\mu-\nu)\,p_t(y)\,\mathrm{d}y \\
&\leq \int\limits_y \left(\sup_v \int v\,\mathrm{d}(\mu-\nu)\right) p_t(y)\,\mathrm{d}y \\
&= \widetilde{\mathrm{W}}_1\left(\mu,\nu\right)
\end{aligned}$$

En passant au sup dans le membre de gauche, on obtient l'inégalité voulue.

Nous allons minorer $\widetilde{\mathrm{W}}_1\left(\mu_{n,t},\nu_{n,t}\right)$ et optimiser en t.

On pose :

$$h(x) = \sum_{m\in\mathbb{Z}^d\setminus\{0\}} \frac{1}{|m|^2} e^{-|m|^2 t}\left(\mathbb{E}\left(e^{i\langle m,X\rangle}|X\sim\mu_n\right) - \mathbb{E}\left(e^{i\langle m,Y\rangle}|Y\sim\nu_n\right)\right) e^{-i\langle m,x\rangle}$$

On introduit des notations :

$$\begin{aligned}
E_\mu(m) &= \mathbb{E}\left(e^{i\langle m,X\rangle}|X\sim\mu_n\right) \\
E_\nu(m) &= \mathbb{E}\left(e^{i\langle m,Y\rangle}|Y\sim\nu_n\right) \\
k_m &= e^{-|m|^2 t}(E_\mu(m)-E_\nu(m))
\end{aligned}$$

On peut donc simplifier, et on a :

$$h(x) = \sum_{m\neq 0} \frac{1}{m^2} k_m e^{-i\langle m,x\rangle}$$

$$\Delta h(x) = \sum_{m\neq 0} k_m e^{-i\langle m,x\rangle}$$

Et pour toute fonction lipschitzienne v :

$$\int v\,\mathrm{d}(\mu_{n,t}-\nu_{n,t}) = -\int v\Delta h\,\mathrm{d}\mathbb{P} = \int \langle \nabla h, \nabla v\rangle\,\mathrm{d}\mathbb{P} \tag{3.4.4}$$

On utilise alors le lemme d'approximation suivant :

Lemme 3.4.5 (Approximation de Luzin)**.** *Il existe $K > 0$ tel que pour tout $\alpha > 0$, et pour toute fonction $h : \mathbb{T}^d \to \mathbb{R}$ $\mathcal{C}^1$ sur le tore, il existe $u : \mathbb{T}^d \to \mathbb{R}$ une fonction α-lipschitzienne telle que : $\mathbb{P}\{h \neq u\} \leq \frac{K}{\alpha^2}\int |\nabla h|^2\,\mathrm{d}\mathbb{P}$*

La démonstration de ce lemme est repoussée à la section 5.

Pour $\alpha > 0$ on pose $u : Q^d \mapsto \mathbb{R}$ la fonction α-lipschitzienne décrite par le lemme précédent, telle que $\mathbb{P}(u \neq h) \leq \frac{K}{\alpha^2}\int |\nabla h|^2\,\mathrm{d}\mathbb{P}$. On a alors :

$$\widetilde{\mathrm{W}}_1\left(\mu_n, \nu_n\right) \geq \widetilde{\mathrm{W}}_1\left(\mu_{n,t}, \nu_{n,t}\right) \geq \frac{1}{\alpha}\left|\int u\,\mathrm{d}(\mu_{n,t}-\nu_{n,t})\right|$$

$$\begin{aligned}\int u\,\mathrm{d}(\mu_{n,t}-\nu_{n,t}) &= \int \langle \nabla h, \nabla u\rangle\,\mathrm{d}\mathbb{P}\\ &= \int |h|^2\,\mathrm{d}\mathbb{P} - \int\limits_E \langle \nabla h, \nabla(h-u)\rangle\,\mathrm{d}\mathbb{P}\end{aligned}$$

Avec $E = \{h \neq u\} \subseteq Q^d$ et $\mathbb{P}(E) \leq \frac{K}{\alpha^2}\int |\nabla h|^2\,\mathrm{d}\mathbb{P}$.

D'où,

$$\alpha\,\mathbb{E}\left(\widetilde{\mathrm{W}}_1\left(\mu_n, \nu_n\right)\right) \geq \mathbb{E}\left(\int |\nabla h|^2\,\mathrm{d}\mathbb{P}\right) - \mathbb{E}\left(\left|\int\limits_E \langle \nabla h, \nabla h - \nabla u\rangle\,\mathrm{d}\mathbb{P}\right|\right) \tag{3.4.6}$$

Comme les X_i et les Y_i suivent des lois uniformes et sont indépendantes, on obtient des simplifications d'espérances : pour des vecteurs m, p, q, r non nuls :

$$\begin{aligned}\mathbb{E}\left(E_\mu(m)\right) &= 0\\ \mathbb{E}\left(E_\mu(m)E_\mu(p)\right) &= \frac{1}{n}\mathbb{1}(m+p=0)\\ \mathbb{E}\left(E_\mu(m)E_\mu(p)E_\mu(q)E_\mu(r)\right) &= \frac{1}{n^3}\mathbb{1}(m+p+q+r=0)\\ &+ \frac{n-1}{n^3}(\mathbb{1}(m+p=q+r=0)\\ &+ \mathbb{1}(m+q=p+r=0) + \mathbb{1}(m+r=p+q=0))\end{aligned}$$

Des égalités similaires sont vraies pour les E_ν et pendant les calculs on gardera en tête que les E_μ sont indépendants des E_ν.

Alors,

$$\begin{aligned}
|\nabla h(x)|^2 &= \sum_{i=1}^{d} \sum_{m,p\neq 0} \frac{1}{|m|^2|p|^2} k_m \overline{k_p} m_i p_i e^{-i\langle m-p,x\rangle} \\
\mathbb{E}\left(k_m\overline{k_p}\right) &= e^{-|m|^2t-|p|^2t}\mathbb{E}\left((E_\mu(m)-E_\nu(m))(E_\mu(-p)-E_\nu(-p))\right) \\
&= e^{-|m|^2t-|p|^2t}\cdot\frac{2}{n}\mathbb{1}(m-p=0) \\
\mathbb{E}\left(|\nabla h(x)|^2\right) &= \frac{2}{n}\sum_{i=1}^{d}\sum_{m\neq 0}\frac{m_i^2}{|m|^4}e^{-2|m|^2t} \\
&= \frac{2}{n}\sum_{m\neq 0}\frac{1}{|m|^2}e^{-2|m|^2t}
\end{aligned}$$

On note :

$$c(n,t) = \mathbb{E}\left(\int |\nabla h|^2 \,\mathrm{d}\mathbb{P}\right) = \frac{2}{n}\sum_{m\neq 0}\frac{1}{|m|^2}e^{-2|m|^2t}$$

D'autre part, d'après les caractéristiques de u, on a :

$$\begin{aligned}
\left|\int_E \langle \nabla h, \nabla h - \nabla u\rangle\right| &\leq \int_E |\nabla h|^2 \,\mathrm{d}\mathbb{P} + \alpha\int_E |\nabla h| \,\mathrm{d}\mathbb{P} \\
&\leq \mathbb{P}(E)^{1/2}\left(\int |\nabla h|^4 \,\mathrm{d}\mathbb{P}\right)^{1/2} + \alpha\mathbb{P}(E)^{3/4}\left(\int |\nabla h|^4 \,\mathrm{d}\mathbb{P}\right)^{1/4} \\
&\leq \left(\frac{K}{\alpha^2}\int |\nabla h|^2 \,\mathrm{d}\mathbb{P}\right)^{1/2}\left(\int |\nabla h|^4 \,\mathrm{d}\mathbb{P}\right)^{1/2} \\
&+ \alpha\left(\frac{K}{\alpha^2}\int |\nabla h|^2 \,\mathrm{d}\mathbb{P}\right)^{3/4}\left(\int |\nabla h|^4 \,\mathrm{d}\mathbb{P}\right)^{1/4}
\end{aligned}$$

Par un calcul similaire à celui de $c(n,t)$, on obtient, pour $d(n,t) = \mathbb{E}\left(\int |\nabla h(x)|^4\right)$:

$$\begin{aligned}
d(n,t) = \frac{2}{n^3} &\sum_{a,b,c,d\neq 0, a+b+c+d=0} \langle a,b\rangle \langle c,d\rangle \frac{e^{-\left(|a|^2+|b|^2+|c|^2+|d|^2\right)t}}{|a|^2|b|^2|c|^2|d|^2} \\
&+ \frac{4n+2}{n^3}\sum_{a,b\neq 0}\langle a,b\rangle^2 \frac{e^{-2(|a|^2+|b|^2)t}}{|a|^4|b|^4}
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{4n-2}{n^3} \sum_{a,b\neq 0} \frac{e^{-2(|a|^2+|b|^2)t}}{|a|^2|b|^2} \\
&\leq \frac{2}{n^3} \sum_{a,b,c,d\neq 0} \frac{e^{-\left(|a|^2+|b|^2+|c|^2+|d|^2\right)t}}{|a||b||c||d|} + \frac{12}{n^2} \sum_{a,b\neq 0} \frac{e^{-2(|a|^2+|b|^2)t}}{|a|^2|b|^2} \\
&= \frac{2}{n^3} \left(\sum_{m\neq 0} \frac{e^{-|m|^2 t}}{|m|} \right)^4 + \frac{12}{n^2} \left(\sum_{m\neq 0} \frac{e^{-2|m|^2 t}}{|m|^2} \right)^2 \\
&= \frac{2}{n^3} \left(\sum_{m\neq 0} \frac{e^{-|m|^2 t}}{|m|} \right)^4 + 3\, c(n,t)^2
\end{aligned}
$$

Par comparaison série-intégrale, on obtient pour $t \to 0$ l'équivalence :

$$
\begin{aligned}
\sum_{m\neq 0} \frac{e^{-|m|^2 t}}{|m|} &\sim \int\limits_{|x|\geq 1} \frac{e^{-|x|^2 t}}{|x|} \,\mathrm{d}x \\
&= \frac{1}{t^{\frac{d-1}{2}}} \int\limits_{|x|\geq \sqrt{t}} \frac{e^{-|x|^2}}{|x|} \,\mathrm{d}x \\
&= \frac{1}{t^{\frac{d-1}{2}}} \int\limits_{r\geq \sqrt{t}} \frac{e^{-r^2}}{r} \sigma_r \,\mathrm{d}r \\
&= \frac{\sigma_1}{t^{\frac{d-1}{2}}} \int\limits_{r\geq \sqrt{t}} e^{-r^2} r^{d-2} \,\mathrm{d}r
\end{aligned}
$$

où $\sigma_r = r^{d-1} \cdot \sigma_1$ est la mesure surface de la sphère de rayon r ; de façon analogue on obtient :

$$
\sum_{m\neq 0} \frac{e^{-2|m|^2 t}}{|m|^2} \sim \frac{\sigma_1}{(2t)^{\frac{d-2}{2}}} \int\limits_{r\geq \sqrt{2t}} e^{-r^2} r^{d-3} \,\mathrm{d}r
$$

À $d = 2$, ces intégrales sont respectivement de l'ordre de (i.e. équivalentes à une constante fois) $\frac{1}{\sqrt{t}}$ et $\log \frac{1}{t}$. Finalement, l'inégalité (3.4.3) se réécrit :

$$
\alpha \mathbb{E} \left(\widetilde{\mathrm{W}}_1 \left(\mu_n, \nu_n\right) \right) \geq c(n,t) - \frac{1}{\alpha} \sqrt{c(n,t)} \sqrt{d(n,t)} - \frac{1}{\sqrt{\alpha}} c(n,t)^{3/4} d(n,t)^{1/4} \tag{3.4.7}
$$

avec $c(n,t) \sim c_1 \cdot \frac{\log 1/t}{n}$ et $d(n,t) \leq \frac{2}{n^3}\frac{c_2}{t^2}(1+o(1)) + 3c(n,t)^2$.

Il ne reste plus qu'à choisir les valeurs de t et de α, sous les seules contraintes $t > 0$, $\alpha > 0$ et $t \to_{n\to\infty} 0$. On pose donc $t_n = \frac{1}{\sqrt{n}}$ et $\alpha_n = \beta \cdot \sqrt{c(n,t_n)}$ et on obtient :

$$\mathbb{E}\left(\widetilde{\mathrm{W}}_1\left(\mu_n,\nu_n\right)\right) \geq \frac{1}{\beta}\sqrt{c(n,t_n)} - \frac{\sqrt{3}}{\beta^2}\sqrt{c(n,t_n)} - \frac{3^{1/4}}{\beta\sqrt{\beta}}\sqrt{c(n,t_n)}$$

On choisit donc $\beta > 0$ suffisamment grand pour avoir $\frac{1}{\beta} - \frac{\sqrt{3}}{\beta^2} - \frac{3^{1/4}}{\beta\sqrt{\beta}} = c > 0$, et on a :

$$\mathbb{E}\left(\widetilde{\mathrm{W}}_1\left(\mu_n,\nu_n\right)\right) \geq c\sqrt{c(n,t_n)} \sim c'\sqrt{\log n/n} \tag{3.4.8}$$

C'est l'ordre de grandeur voulu.

3.4.3 En dimension 1

Pour $d = 1$, on peut écrire explicitement la solution du problème d'appariement. Notons $X^{(1)} \leq X^{(2)} \leq \cdots \leq X^{(n)}$ les variables $X_1, X_2, \ldots, X_n$ ordonnées. La solution optimale au problème d'appariement consiste toujours à relier $X^{(k)}$ à $Y^{(k)}$ pour $k \in \{1,\ldots,n\}$ (cela se déduit de l'inégalité triangulaire).

Ainsi,

$$\begin{aligned}\mathrm{W}_1\left(\mu_n,\nu_n\right) &= \frac{1}{n}\sum_{k=1}^{n}|X^{(k)} - Y^{(k)}| \\ &= \frac{1}{n}\int_0^1 \sum_{k=1}^{n} \mathbb{1}(x \in [X^{(k)}, Y^{(k)}])\,\mathrm{d}x \end{aligned} \tag{3.4.9}$$

Or,à x fixé,

$$\begin{aligned}\sum_{k=1}^{n} \mathbb{1}(x \in [X^{(k)}, Y^{(k)}]) &= |\#\{k|X^{(k)} \leq x\} - \#\{k|Y^{(k)} \leq x\}| \\ &= \left|\sum_{k=1}^{n} \mathbb{1}(X^{(k)} \leq x) - \mathbb{1}(Y^{(k)} \leq x)\right| \end{aligned} \tag{3.4.10}$$

On peut donc écrire :

$$\begin{aligned}
\mathbb{E}\left(\mathrm{W}_1\left(\mu_n,\nu_n\right)\right) &= \frac{1}{n}\mathbb{E}\left(\int_0^1\left|\sum_{k=1}^n \mathbb{1}(X^{(k)}\le x)-\mathbb{1}(Y^{(k)}\le x)\right|\mathrm{d}x\right)\\
&\ge \frac{1}{n}\mathbb{E}\left(\left|\int_0^1\left(\sum_{k=1}^n \mathbb{1}(X^{(k)}\le x)-\mathbb{1}(Y^{(k)}\le x)\right)\mathrm{d}x\right|\right)\\
&= \frac{1}{\sqrt{n}}\mathbb{E}\left(\frac{1}{\sqrt{n}}\left|\sum_{k=1}^n(X_k-Y_k)\right|\right) \qquad (3.4.11)
\end{aligned}$$

Or les $X_k - Y_k$ sont i.i.d et centrées d'écart type $\sigma > 0$. Ainsi, par le théorème de la limite centrale $Z_n = \frac{1}{\sqrt{n}}\sum_{k=1}^n(X_k-Y_k)$ converge en loi vers la loi normale centrée d'écart type σ, $\mathcal{N}$. Il faut alors vérifier que $E(|Z_n|)$ converge vers $E(|\mathcal{N}|)$. Ce n'est pas a priori une conséquence directe du théorème centrale limite.

Cependant, comme $x = \underset{x\to+\infty}{o}(x^2)$, pour tout $\varepsilon > 0$, pour M assez grand, $E(|Z_n|\mathbb{1}_{|Z_n|>M}) \le \varepsilon E(Z_n^2) = \epsilon\sigma^2$. L'espérance de la queue de distribution est donc majorée uniformément en n et la convergence en loi garantit que à $M > 0$ fixé, $\mathbb{E}\left(|Z_n|\mathbb{1}_{|Z_n|\le M}\right)$ converge vers $\mathbb{E}\left(|\mathcal{N}|\mathbb{1}_{|\mathcal{N}|\le M}\right)$ et donc la convergence de $E(|Z_n|)$

Ainsi, il existe une constante $C > 0$ telle que pour n assez grand,

$$\mathbb{E}\left(\mathrm{W}_1\left(\mu_n,\nu_n\right)\right) \ge \frac{C}{\sqrt{n}}. \qquad (3.4.12)$$

4 Théorème de Kantorovitch-Rubinstein

Nous présentons une preuve du théorème de Kantorovitch-Rubinstein (théorème 3.1.1). Cette preuve est à peu de choses près celle présentée dans [8] (chap. 1 §.1 et §.2, pp 17 -34) Pour prouver le théorème, nous avons besoin d'un résultat assez général d'analyse fonctionnelle : le théorème de dualité de Fenchel-Rockafellar.

4.1 Dualité de Fenchel-Rockafellar

Dans cette section, nous énonçons et démontrons le théorème de dualité de Fenchel-Rockafellar.

Théorème 4.1.2 (Dualité de Fenchel-Rockafellar). *Soit E un espace vectoriel topologique. On note E' son dual analytique. Si $\Phi : E \to \mathbb{R}^+ \cup \{+\infty\}$, on définit sa transformée de Fenchel-Legendre*[2] $\Phi^* : E' \to \mathbb{R}^+ \cup \{+\infty\}$ *de la manière suivante :*

$$\Phi^*(f) = \sup_{z \in E}(f(z) - \Phi(z)) \tag{4.1.1}$$

Soit Φ et Ψ deux fonctions convexes. On suppose que Φ est continue en au moins un point de E. Alors,

$$\inf_{x \in E} (\Phi(x) + \Psi(x)) = \max_{f \in E'} (-\Phi^*(-f) - \Psi^*(f)) \tag{4.1.2}$$

Démonstration. On peut établir ce théorème à l'aide du théorème de séparation de Hahn-Banach. Posons $m = \inf_{x \in E} (\Phi(x) + \Psi(x))$ L'inégalité $m \leq \sup_{f \in E'} -\Phi^*(-f) - \Psi^*(f)$ découle de la définition de la transformée de Fenchel-Legendre.

Pour montrer l'autre inégalité, considérons, le convexe $C = \{(x, \lambda) \in E \times \mathbb{R} \,|\lambda \geq \phi(x)\}$ et $C' = \{(y, \mu) \in E \times \mathbb{R}|\mu \geq m - \psi(x)\}$. C est d'intérieur non vide (car ϕ est continue en au moins un point) et $\text{adh}(\text{int}(C)) = C$ car C est convexe (voir par exemple au début du premier chapitre de [5] pour une démonstration de ce fait). Par définition de C et C', $\text{int}(C) \cap C' = \emptyset$. Ainsi, le théorème de séparation de Hahn-Banach garantit l'existence d'une forme linéaire non nulle, $l = (f, \alpha) \in (E \times \mathbb{R})' = E' \times \mathbb{R}$ tel que $\sup_{C'} l < \inf_{\text{int}(C)} l$ et donc $\sup_{C'} l \leq \inf_{C} l$. Ainsi, il existe $f \in E'$ et $\alpha \in \mathbb{R}$ tel que pour tout $(x, \lambda) \in C$ et pour tout $(y, \mu) \in C'$,

$$f(x) + \alpha\lambda \geq f(y) + \alpha\mu.$$

Faisant tendre λ vers l'infini, on voit que $\alpha \geq 0$. Si α est nul, on obtient $\forall x, y \in E$, $f(x) = f(y)$, autrement dit, f est nulle ce qui est exclus car l est non nulle. Ainsi,

$$\forall x, y \in E, \frac{f}{\alpha}(x) + \lambda \geq \frac{f}{\alpha}(y) + m - \psi(y) \text{ d'où } -\phi^*(-\frac{f}{\alpha}) - \psi^*(\frac{f}{\alpha}) \geq m. \tag{4.1.3}$$

Il y a donc égalité dans (4.1.2) en $\frac{f}{\alpha}$. □

4.2 Dualité de Kantorovitch

La dualité de Kantorovitch est un résultat général de la théorie du transport optimal.

2. Le premier chapitre de [5] fait une présentation plus détaillée de la fonction maximale et propose une preuve légèrement différente du théorème

Théorème 4.2.1. *Soit X et Y deux espaces polonais (séparable et complet) munis de leur tribus boreliennes. Soit μ (respectivement ν) deux mesures de probabilités sur X (respectivement Y). Soit $c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ une fonction semi-continue inférieurement. On note $\Phi_c = \left\{(\varphi, \psi) \in \mathrm{L}^1(\mathrm{d}\mu) \times \mathrm{L}^1(\mathrm{d}\nu) | \varphi(x) + \psi(y) \leq c(x,y)\, \mu \otimes \nu \text{ p.p}\right\}$. Alors,*

$$\inf_{\pi \in \Pi(\mu,\nu)} \int c(x,y)\,\mathrm{d}\pi(x,y) = \sup_{(\varphi,\psi)\in\Phi_c} \int_X \varphi(x)\,\mathrm{d}\mu(x) + \int_Y \psi(y)\,\mathrm{d}\nu(y) \tag{4.2.2}$$

Cafarelli donne une explication amusante et intuitive du théorème rapportée par Villani dans [8, §1.1.3]) et librement traduite et adaptée ici. Imaginez que vous êtes un vendeur de fraises qui souhaite acheminer des produits provenant de diverses exploitations agricoles vers différents magasins. Vous vous apprêtez à louer des camions et à payer un côut $c(x, y)$ par kilogramme de fraises transportées entre x et y. Votre ami mathématicien vous propose alors le marché suivant : il prend en charge entièrement l'acheminement et vous fait payer un prix $\varphi(x)$ pour chaque kilogramme de fruits dans l'exploitation x et un prix $\psi(y)$ pour chaque kilo de fraisesà apporter dans le magasin y. Il vous garantit que le coût de transport de chaque fraise sera moindre que ce que vous auriez payé en louant des camions (quitte à fixer des prix négatifs pour compenser quand c'est nécessaire). Vous acceptez le marché qui est bien sûr avantageux. La dualité de Kantorovitch garantit à votre ami qu'il peut vous faire payer (presque) autant que le loueur de camions s'il arrange bien ses prix.

Démonstration. On se restreint au cas où X et Y sont des espaces métriques compacts et c continu (c'est le seul que nous utilisons). Il n'est pas difficile de voir qu'il suffit de prouver l'égalité en restreignant le sup dans le terme de droite aux fonctions continues bornées.

Le théorème de Riesz sur le dual des fonctions continues sur un compact garantit que $\mathcal{M}(X \times Y)$ est le dual analytique des fonctions continues sur $X \times Y$. Le théorème s'obtient en appliquant la dualité de Fenchel-Rockafeller aux fonctions de $\mathcal{C}(X \times Y)$ dans $\mathbb{R}$ suivantes :

$$\Phi : u \mapsto \begin{cases} 0 \text{ si } \forall x, y \in X \times Y, u(x,y) \geq -c(x,y) \\ +\infty \text{ sinon} \end{cases} \tag{4.2.3}$$

(continue en $u = 1$) et :

$$\Psi : u \mapsto \begin{cases} \int_X f\,\mathrm{d}\mu + \int_Y g\,\mathrm{d}\nu \text{ si } \forall x, y \in X \times Y, u(x,y) = f(x) + g(y) \\ +\infty \text{ sinon} \end{cases} \tag{4.2.4}$$

On calcule donc pour $\pi \in M(X \times Y)$,

$$\Phi^*(-\pi) = - \inf_{\substack{u \in E, \\ \forall x \in X, \forall y \in Y, \\ u(x,y) \geq -c(x,y)}} \int_{X \times Y} u \,\mathrm{d}\pi$$

$$= \begin{cases} \int_{X \times Y} c \,\mathrm{d}\pi \ \text{ si } \pi \in \mathcal{M}_+(X \times Y) \\ +\infty \text{ sinon} \end{cases} \tag{4.2.5}$$

En effet, si π n'est pas une mesure positive, il existe $u \geq 0 \geq -c$ telle que $\int_{X \times Y} u \,\mathrm{d}\pi < 0$. En multipliant un tel u par une constante arbitrairement grande, l'inf ci-dessus est infini.

De la même manière,

$$\Psi^*(\pi) = \begin{cases} 0 \text{ si } \pi \in \Pi(\mu, \nu) \\ +\infty \text{ si } \pi \in \Pi(\mu, \nu) \end{cases} \tag{4.2.6}$$

Ainsi, pour $\pi \in \mathcal{M}(X \times Y)$,

$$-\Phi^*(-\pi) - \Psi^*(\pi) = \begin{cases} -\int c \,\mathrm{d}\pi \ \text{ si } \pi \in \Pi(\mu, \nu) \\ -\infty \text{ sinon} \end{cases} \tag{4.2.7}$$

Et si $u \in \mathcal{C}(X \times Y)$,

$$\Phi(u) + \Psi(u) = - \sup_{(f,g) \in \Phi_c \cap \mathcal{C}(X \times Y} \int \phi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu \tag{4.2.8}$$

Ce qui achève la preuve. $\square$

4.3 Théorème de Kantorovitch-Rubinstein

On peut finalement démontrer le théorème de Kantorovitch-Rubinstein (théorème 3.1.1).

L'inégalité

$$\mathrm{W}_1(\mu, \nu) \geq \sup_{\|f\|_{\mathrm{Lip}}} \int_X f \,\mathrm{d}(\mu - \nu) \tag{4.3.1}$$

découle immédiatement du caractère lipschitzien de f. L'autre inégalité se déduit du théorème de dualité de Kantorovitch. En effet, lorsque la fonction de coût est une distance, on peut restreindre le sup dans le théorème de Kantorovitch aux seuls couples de fonctions de la forme $(f, -f)$ avec f 1- lipschitzienne. Si $f : X \to \mathbb{R}$, on introduit la fonction c-concave conjuguée, $f^c : X \to \mathbb{R} \cup \{-\infty\}$ définie par :

$$f^c(x) = \inf_{y \in E} d(x, y) - f(y). \tag{4.3.2}$$

Cette fonction est 1-lipschitzienne car c'est un infimum de fonction 1-lipschitzienne [3]. De plus si f est 1-lipschitzienne, alors $f^c = -f$. Pour $(\phi, \psi) \in \Phi_d$, $\forall x, y \in X, \phi(y) \leq d(x, y) - \psi(x)$ donc $\psi(y) \leq \phi^c(y)$.
Ainsi,

$$\begin{aligned}\int_X \phi(x)\,\mathrm{d}\mu(x) + \int_X \psi(y)\,\mathrm{d}\nu(y) &\leq \int_X \phi(x)\,\mathrm{d}\mu(x) + \int_X \phi^c(y)\,\mathrm{d}\nu(y) \\ &\leq \int_X (\phi^c)^c(x)\,\mathrm{d}\mu(x) + \int_X \phi^c(y)\,\mathrm{d}\nu(y) \\ &= \int_X \phi^c(y)\,\mathrm{d}(\nu - \mu)(y) \qquad (4.3.3)\end{aligned}$$

5 Approximation de Luzin et fonctions maximales

Dans cette section, on présente une démonstration du lemme 3.4.5 d'approximation par des fonctions lipschitiennes. La preuve repose sur l'utilisation de la fonction maximale. Une preuve détaillée dans un ouvert suffisamment régulier de $\mathbb{R}^d$ est présentée dans [1]. Nous suivons cette preuve en adaptant les raisonnements au tore. Soit $f : \mathbb{T}^d \to \mathbb{R}$ une fonction $\mathcal{C}^\infty$ définie sur le tore euclidien $\mathbb{T}^d$ (dans cette section uniquement, le tore est vu comme $\mathbb{R}^d/\mathbb{Z}^d$ par commodité). Le tore est naturellement muni d'une structure géométrique rigide où les changements de cartes sont donnés par des translations. Ceci permet de définir le gradient de f directement comme une fonction à valeur dans $\mathbb{R}^d$ et non dans le fibré tangent du tore. On définit alors la fonction maximale de f par :

$$M(f)(x) = \sup_{r \in \mathbb{R}_+} \int_{B_{x,r}} h(t)\,\mathrm{d}t \tag{5.1}$$

Nous aurons besoin des deux propriétés suivantes de la fonction maximale :

3. à y fixé, la fonction dans l'infimum est la distance translatée d'une constante.

Lemme 5.2. *Il existe $C > 0$ ne dépendant que de d (indépendant de α) tel que quel que pour toute fonction $\mathcal{C}^\infty$ f et $\forall x, y \in \{z|\ |M(\nabla f)(z)| < \alpha\}$*

$$\left|\frac{f(x) - f(y)}{D(x, y)}\right| < C\alpha$$

Démonstration. Une preuve dans un ouvert suffisamment régulier de $\mathbb{R}^d$ est détaillée dans [1]. Nous reprenons cette preuve et l'adaptons au cas du tore en évitant l'utilisation des théorèmes d'extension dans les espaces de Sobolev. On note D la distance euclidienne sur le tore et μ la mesure de Lebesgue sur le tore. Pour $\alpha > 0$, soit $E_\alpha = \{z|\,|M(\nabla f)(z)| < \alpha\}$. Soit $x, y \in E_\alpha$. On note $r = D(x, y)$ Posons pour $k > 0$,

$$A_k = \{z \in B(x, r)|\,|\nabla f)(z)| > k\alpha\} \quad \text{et} \quad B_k = \{z \in B(y, r)|\,|\nabla f)(z)| > k\alpha\}.$$

On contrôle la mesure de A_k (et de B_k) par l'inégalité suivante :

$$\mu(A_k) \leq \frac{K\mu(B(x, r))}{k} \tag{5.3}$$

où $K > 0$ ne dépend que de d. Soit H l'hypercube semi-ouvert $\tilde{x} + [0, 1[^d$ où $\tilde{x}$ est un représentant de x dans R^d. Si $z \in \mathbb{T}^d$, on note $\tilde{z}$, l'unique représentant de z dans H. On pose $\tilde{f} : H \to \mathbb{R}$ définie par $\tilde{f}(\tilde{z}) = f(z)$. Pour tout point $\tilde{z}$ de H privé de son bord, $\nabla_{\tilde{z}}\tilde{f} = \nabla_z f$. Pour tout $z \in \mathbb{T}^d$, $D(z, x) = |\tilde{z} - \tilde{x}|$. On notera également f la fonction $\tilde{f}$ dans la suite. On note $B(x, r)$ (resp. $B_{\mathbb{R}^d}(x, r)$) la boule de centre x et de rayon r dans $\mathbb{T}^d$ (resp $\mathbb{R}^d$).

Par l'inégalité de Markov :

$$\mu(A_k) \leq \frac{1}{k\alpha} \int\limits_{B(x,r)\subseteq Q^d} \frac{|f(z) - f(x)|}{D(z, x)}\,\mathrm{d}z \tag{5.4}$$

Or, pour $z \in \mathbb{T}^d$, $|f(z) - f(x)| = |\int_0^1 \left\langle \nabla_{x+t(\tilde{z}-\tilde{x})} f, \tilde{z} - \tilde{x} \right\rangle \mathrm{d}t\,| \leq \int_0^1 \left|\nabla_{\tilde{x}+t(\tilde{z}-\tilde{x})} f\right| D(x, z)\,\mathrm{d}t$ Ainsi,

$$\mu(A_k) \leq \frac{1}{k\alpha} \int\limits_{B(x,r)\subseteq Q^d} \int\limits_0^1 \left|\nabla_{\tilde{x}+t(\tilde{z}-\tilde{x})} f\right| \mathrm{d}t\,\mathrm{d}z \tag{5.5}$$

$$= \int\limits_0^1 \int\limits_{B(x,r)\subseteq Q^d} \left|\nabla_{\tilde{x}+t(\tilde{z}-\tilde{x})} f\right| \mathrm{d}t\,\mathrm{d}z \tag{5.6}$$

En faisant le changement de variable $u = t(\tilde{z} - \tilde{x})$ à t fixé, comme r est inférieur au diamètre du tore (qui est aussi le diamètre de H), on est ramené à intégrer sur l'image par l'homotétie de centre x et de facteur t de $B_{\mathbb{R}^d}(\tilde{x}, r) \cap H$. Cet ensemble est inclus dans $B_{\mathbb{R}^d}(x, tr) \cap H$ Ainsi,

$$\begin{aligned}
\mu(A_k) &\leq \frac{1}{k\alpha} \int_0^1 \frac{1}{t^d} \int_{B_{\mathbb{R}^d}(x,tr)\cap H} |\nabla_u f| \, \mathrm{d}u \, \mathrm{d}t && (5.7)\\
&\leq \int_0^1 \frac{1}{k\alpha} \frac{\mu(B(x,tr))}{t^d} M(|\nabla f|)(x) \, \mathrm{d}t && (5.8)\\
&\leq \frac{\mu(B_{\mathbb{R}^d}(x,r))}{k\alpha} M(|\nabla f|)(x) \qquad \text{car } B_{\mathbb{R}^d}(x,tr) \cap H \subseteq B_{\mathbb{R}^d}(x,tr) && (5.9)\\
&\leq \frac{K}{k} \mu(B(x,r)) && (5.10)
\end{aligned}$$

où $K = \sup_{r\in[0,1]} \frac{\mu(B_{\mathbb{R}^d}(x,r))}{\mu(B(x,r))} < \infty$.

Si $x, y \in \mathbb{T}^d$ sont tels que $D(x,y) = r$, alors $\mu(B(x,r) \cap B(y,r)) > L\mu B(x,r)$ où $L > 0$ est une constante indépendante de r. En effet, pour r assez petit, $B(x,r) \cup B(y,r)$ est isométrique $B_{\mathbb{R}^d}(x,r) \cup B_{\mathbb{R}^d}(y,r)$. Pour

$$k > \frac{1}{2KL},\ \mu(A_k \cup B_k) < \mu(B(x,r) \cap B(y,r)).$$

En particulier, il existe $z \in (\mu(B(x,r) \cap B(y,r)) \setminus (A_k \cup B_k)$.
Ainsi,

$$\frac{|f(x) - f(y)|}{D(x,y)} \leq \frac{|f(x) - f(z)|}{D(x,z)} + \frac{|f(y) - f(z)|}{D(y,z)} \leq 4LK\alpha \tag{5.11}$$

□

Lemme 5.12. *Soit f une fonction continue.*

$$\mu(M(f) > \alpha) \leq \frac{5^d}{\alpha^2} \int_{\mathbb{T}^d} |f|^2$$

où μ désigne ici la mesure de Lebesgue.

Démonstration. La démonstration repose essentiellement sur le lemme de recouvrement de Vitali. Notons d'abord que $M(f)^2 \leq M(f^2)$ par convexité de la fonction carré. Il suffit donc de montrer que pour toute fonction f, $\mu(M(f) > \alpha) \leq \frac{5^d}{\alpha} \int_{\mathbb{T}^d} |f|$ et de l'appliquer à f^2.

Pour tout $x \in (M(f) > \alpha)$, il existe $r > 0$ tel que $\int_{B(x,r)} |f| \geq \mu(B(x,r))\alpha$. Fixons un tel r_x pour chaque x. Par le lemme de Vitali, on peut extraire une sous-famille au plus dénombrable $(x_n)_{n\in I} \subseteq (M(f) > \alpha)$ telles que les $B(x_n, r_n)$ soient deux à deux disjointes et $(M(f) > \alpha) \subseteq \cup_{n\in I} B(x, 5r_x)$. Ainsi,

$$\mu(M(f) > \alpha) \leq \sum_{n\in I} \mu(B(x, 5r_x)) \leq 5^d \sum_{n\in I} \mu(B(x, r_x)) \leq \frac{5^d}{\alpha} \int_{\mathbb{T}^d} |f|. \tag{5.13}$$

□

Pour conclure, on prolonge la restriction de f à $(M(|\nabla(f)| < \alpha)$ à $\mathbb{T}^d$ tout entier en une fonction u qui est $C\alpha$ lipschitzienne. Ceci est permis par le lemme suivant :

Lemme 5.14. *Soit X un espace métrique et $A \subseteq X$. Soit $f : A \to \mathbb{R}$ une fonction k-lipschitzienne sur A. Alors, il existe un prolongement k-lipschitz $u : X \to \mathbb{R}$.*

Démonstration. $u(x) = \inf_{a\in A}(f(a) + kd(x,a))$ convient (par exemple). □

Enfin, u est automatiquement faiblement dérivable et $\|\nabla u\|_\infty \leq C\alpha$. (voir le théorème 4 de [7, §5.8.1]).

Références

[1] Emilio Acerbi et Nicola Fusco : An approximation lemma for $W^{1,p}$ functions. Material instabilities in continuum mechanics, Proc. Symp. Edinburgh/Scotl. 1985/86, 1-5 (1988)., 1988.

[2] M. Ajtai, J. Komlós et Gábor Tusnády : On optimal matchings. *Combinatorica*, 4:259–264, 1984.

[3] Luigi Ambrosio, Federico Stra et Dario Trevisan : A PDE approach to a 2-dimensional matching problem. *Probab. Theory Relat. Fields*, 173(1-2):433–477, 2019.

[4] Sergey G. Bobkov et Michel Ledoux : A simple Fourier analytic proof of the AKT optimal matching theorem. *Ann. Appl. Probab.*, 31(6):2567–2584, 2021.

[5] Haim Brezis : *Functional analysis, Sobolev spaces and partial differential equations*, pages 1–30. Universitext. New York, NY : Springer, 2011.

[6] S. Caracciolo, C. Lucibello, G. Parisi et G. Sicuro : Scaling hypothesis for the euclidean bipartite matching problem. *Physical Review E*, 90(1), jul 2014.

[7] Lawrence C. Evans : *Partial differential equations*, volume 19 de *Grad. Stud. Math.*, pages 294–295. Providence, RI : American Mathematical Society (AMS), 2nd ed. édition, 2010.

[8] Cédric Villani : *Topics in optimal transportation*, volume 58 de *Grad. Stud. Math.*, pages 17–38. Providence, RI : American Mathematical Society (AMS), 2003.