

# Mémoire de stage "Le théorème AKT"

Vincent Viau, sous la direction de Djalil Chafaï

3 septembre 2021

## Table des matières

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Résultats principaux et avancées récentes</b>                   | <b>3</b>  |
| 1.1      | Théorème originel de Ajtai, Komlós et Tusnády . . . . .            | 3         |
| 1.2      | Extension du résultat pour toute loi à support compact . . . . .   | 4         |
| 1.3      | Constante optimale en dimension 1 et 2 : cas uniforme . . . . .    | 7         |
| 1.4      | Constante optimale en dimension 3 et plus : cas uniforme . . . . . | 11        |
| 1.5      | Cas gaussien . . . . .   | 13        |
| <b>2</b> | <b>Quelques avancées personnelles</b>                              | <b>15</b> |
| 2.1      | Borne supérieure pour des lois générales . . . . .                 | 16        |
| 2.2      | Borne inférieure par comparaison avec l'uniforme . . . . .         | 19        |
| 2.3      | Conjecture dans le cas gaussien . . . . .                          | 20        |
| <b>3</b> | <b>Questions ouvertes et problèmes analogues</b>                   | <b>26</b> |
| <b>4</b> | <b>Appendice</b>   | <b>28</b> |
| 4.1      | Théorème de Birkhoff-Von Neumann . . . . .                         | 28        |
| 4.2      | Bases sur le transport optimal . . . . .                           | 29        |
| 4.3      | Bases sur les semi-groupes de Markov . . . . .                     | 31        |
| 4.4      | Bases sur les processus de Poisson . . . . .                       | 34        |
|          | <b>Références</b>  | <b>35</b> |

## Introduction et remerciements

Ce mémoire est le résultat d'un stage de 5 mois effectué de février à juin 2021, sous la direction de Djalil Chafaï. Le sujet initial de ce stage était l'étude de l'article [BL19b] de S. Bobkov et M. Ledoux, mais le stage a finalement plus été une étude générale du théorème AKT, de ses variantes, et de ses prolongements. Je tiens tout d'abord à remercier chaleureusement Djalil Chafaï pour sa disponibilité, surtout en ces temps particuliers, et sa patience. Cela a été une vraie chance de pouvoir échanger aussi régulièrement avec vous, et cela m'aura été, je pense, particulièrement instructif.

Le théorème AKT tient son nom de M. Ajtai, J. Komlós et G. Tusnády, qui sont les premiers à l'avoir démontré en 1984 dans [AKT84]. Ce théorème est un résultat de matching aléatoire, un domaine des probabilités qui s'étend bien au delà de ce simple problème. Commençons par présenter le problème de manière informelle.

Le problème est le suivant : nous considérons deux fois  $n$  tirages indépendants et uniformément distribués sur l'hypercube  $[0, 1]^d$ , où  $d \geq 1$ , que l'on peut représenter par des variables aléatoires  $X_i$  et  $Y_i$ ,  $i = 1, \dots, n$  iid de loi uniforme sur l'hypercube. Nous pouvons alors nous

poser la question assez naturelle suivante : quelle va être la distance entre l'échantillon composé des  $X_i$  et celui composé des  $Y_i$  quand  $n$  devient très grand. Il faut bien sûr préciser cela en définissant la distance entre les deux échantillons. Pour cela, disons que les  $X_i$  sont des points rouges et les  $Y_i$  des points bleus. Nous relierons chaque point rouge à un point bleu, de manière bijective, et nous regardons la somme des distances entre chaque paire de points. Puis nous nous intéressons finalement au minimum que nous pouvons obtenir en considérant toutes les paires possibles entre points bleus et points rouges. L'objectif va alors être de comprendre le comportement de cette valeur quand le nombre de points devient très grand (tend vers l'infini).

Plus formellement, nous nous intéressons à la valeur de  $\inf_{\pi} \frac{1}{n} \sum_{i=0}^n d(X_i, Y_{\pi(i)})$  où l'infimum est pris sur l'ensemble des permutations  $\pi$  de  $\{1, \dots, n\}$  et où  $d$  est une distance sur  $[0, 1]^d$ . Les cas qui ont été les plus étudiés, dont le théorème AKT fait partie, sont les cas où  $d$  provient d'une norme  $\|\cdot\|_p$ , avec  $1 \leq p < +\infty$ . En particulier le théorème AKT original démontré dans [AKT84] s'intéresse au cas où la distance provient de la norme euclidienne, c'est-à-dire au cas  $p = 2$ .

Commençons par donner une heuristique afin d'avoir une idée de ce que pourrait être une borne supérieure. Si nous avons  $n$  points  $X_i$  et  $n$  points  $Y_i$ , étant donné un tirage  $X_i$  nous pouvons nous attendre à ce que le point  $Y_j$  le plus proche se trouve dans un espace de volume de l'ordre de  $\frac{1}{n}$ , et donc  $\|X_i - Y_j\|_p$  devrait être de l'ordre de  $n^{-\frac{p}{d}}$ . Supposons que chaque  $X_i$  soit associé à un  $Y_i$  de cette sorte, il vient alors qu'une borne supérieure devrait être  $O(n^{-\frac{p}{d}})$ . Ce raisonnement très simple est en fait diablement efficace, au moins en grande dimension, puisque cette asymptotique se trouve être la bonne en dimension  $d \geq 3$ . En revanche, en dimension 1 et 2 ce raisonnement ne donne plus la bonne asymptotique. Le cas de la dimension 1 est très particulier, puisqu'il s'obtient grâce à des réarrangements croissants et donne l'asymptotique  $n^{-\frac{p}{d}}$ . Résoudre le problème en dimension 1 est significativement plus simple que pour les autres dimensions, et il existe même des résultats exacts dans le cas  $p = 2$ . Par ailleurs, le cas de la dimension 1 a été grandement étudié, notamment dans [BL19a]. C'est le cas de la dimension 2 qui est en fait le plus intéressant et le plus surprenant. En effet, dans leur article original M. Ajtai, J. Komlós et G. Tuszányi ont montré que l'asymptotique de la grandeur considérée était  $(\frac{\log n}{n})^{\frac{p}{2}}$ . Il y a donc réellement un phénomène de petite dimension dans ce problème de matching, qui peut s'expliquer par la présence, dans des petites régions de l'hypercube, de différences entre le nombre de  $X_i$  et le nombre de  $Y_i$  présents dans cette région, et qui crée donc des phénomènes "d'association longue-distance" entre les tirages.

Il y a plusieurs manières d'aborder le problème. Les arguments utilisés originellement dans [AKT84] ou encore ceux utilisés par M. Talagrand dans [Tal92] relèvent plus de la combinatoire, et semblent peut-être être plus de nature probabiliste. Les arguments utilisés notamment par L. Ambrosio, F. Stra et D. Trevisan dans [AST19] ou ceux utilisés par S. Bobkov et M. Ledoux dans [BL19b] ou M. Ledoux dans [Led19a] peuvent sans doute être qualifiés de plus analytiques, puisqu'ils font appel à des outils comme les semi-groupes de la chaleur, ou encore le transport optimal. Nous donnons dans ce mémoire, avec des degrés de précision variables, les grandes lignes des preuves des résultats importants.

La première partie est une sorte de panorama de l'état actuel de la recherche sur ce problème de matching aléatoire et sur ses différentes variantes et extensions que nous présenterons. Nous essayons pour chaque résultat de donner une idée de la démonstration, au moins dans les grandes lignes, ce qui permet notamment de comparer les différentes méthodes utilisées au fil du temps.

La seconde partie est dédiée à des résultats personnels, résultats des recherches de ce stage. Nous présentons notamment une borne supérieure pour des lois générales, et quelques résultats dans le cas où les tirages suivent des lois gaussiennes.

La troisième partie est consacrée à des questions directement ou indirectement en lien avec le théorème AKT et qui sont aujourd'hui encore ouvertes. Nous présentons notamment les différentes conjectures qui existent à ce jour.

Enfin la dernière partie est un appendice consacrée à des résultats préliminaires généraux, dont le lien avec le problème de matching est plus ou moins évident, mais dont nous avons besoin pour nous attaquer au problème. Nous rappelons un théorème de Birkhoff-Von Neumann sur les points extrémaux des matrices bistochastiques. Ce résultat nous permet d'exprimer  $\inf_{\pi} \frac{1}{n} \sum_{i=0}^n \|X_i, Y_{\pi(i)}\|_p$  comme une distance de Wasserstein entre deux mesures empiriques, et nous avons donc besoin de donner quelques résultats basiques de transport optimal. Ensuite, nous effectuons quelques rappels sur les semi-groupes de Markov, et sur les semi-groupes de la chaleur, qui sont par exemple très utiles dans [AST19]. Nous effectuons enfin une brève présentation des processus de Poisson, qui sont présents dans certaines démonstrations que nous avons rencontrées. Nous ne ferons dans cette partie que très peu de démonstrations.

## 1 Résultats principaux et avancées récentes

### 1.1 Théorème original de Ajtai, Komlós et Tusnády

L'article original de M. Ajtai, J. Komlós et G. Tusnády, concerne uniquement la dimension 2, les autres dimensions avaient déjà été traitées plus tôt, soit par des arguments de réarrangements croissants en dimension 1, soit par des arguments de sous-additivité en dimension supérieure à 3. Comme présenté en introduction, la dimension 2 est vraiment spécifique pour ce problème, comme nous le constatons avec l'apparition du facteur logarithmique.

Remarquons que, par souci d'anachronisme, nous reprenons ici l'énoncé original utilisé dans leur article par les trois auteurs, celui-ci étant équivalent à la manière "Transport Optimal" d'énoncer le problème.

**Théorème 1.1** (Théorème AKT). *Si  $X_i, Y_i$  sont iid de loi uniforme sur le carré unité  $[0, 1]^2$ , et si  $T_n(p) := \min_{\pi} \sum_{i=0}^n |X_i - Y_{\pi(i)}|^p$  où l'infimum est pris sur l'ensemble des permutations  $\pi$  de  $\{1, \dots, n\}$ . Alors, avec probabilité  $1 - o(1)$  :*

$$\frac{1}{C} \left( \frac{\log n}{n} \right)^{\frac{p}{2}} \leq \mathbb{E} \left( \frac{T_n(p)}{n} \right) \leq C \left( \frac{\log n}{n} \right)^{\frac{p}{2}},$$

où  $C > 0$  est une constante dépendant uniquement de  $p$ .

Nous donnons ici une brève idée de la preuve originale pour la borne inférieure et dans le cas où  $p = 1$ , donnée par les auteurs dans [AKT84].

*Idee de preuve de la borne inférieure.* L'idée générale pour montrer la borne inférieure est de construire une fonction  $L$ -Lipschitz  $f$  sur  $Q := [0, 1]^2$  telle que

$$A(f) := \sum_{i=1}^n f(X_i) - \sum_{i=1}^n f(Y_i) \geq LC_1(n \log n)^{\frac{1}{2}}, \quad (1.1)$$

et de se servir ensuite de la dualité de Kantorovich (4.2) pour conclure.

Donnons d'abord une idée de ce à quoi la fonction  $f$  en question devrait ressembler, bien que cette première approche échoue puisque la fonction que nous exhibons dans un premier temps n'est pas Lipschitz. Soit  $m = \frac{\log n}{10}$ . Pour  $i = 1, \dots, m$  nous notons  $Q_{i,j}$ , pour  $j = 1, \dots, 4^i$  les sous-carrés de  $Q$  obtenus en divisant les deux coordonnées en  $2^i$  parts égales. Nous définissons ensuite

$$X(Q_{i,j}) := \sum_k \mathbb{1}_{Q_{i,j}}(X_k), \quad Y(Q_{i,j}) := \sum_k \mathbb{1}_{Q_{i,j}}(Y_k), \quad V(Q_{i,j}) := X(Q_{i,j}) - Y(Q_{i,j}). \quad (1.2)$$

puis, pour la fonction  $f(x) := \sum_{i,j} V(Q_{i,j}) \mathbb{1}_{Q_{i,j}}(x)$ , nous avons

$$A(f) = \sum_{i,j} V^2(Q_{i,j}) \sim Cn \log n. \quad (1.3)$$

Malheureusement, comme annoncé juste avant, la fonction  $f$  ainsi obtenue n'est pas Lipschitz. Il s'agit donc maintenant de redéfinir  $V$  et  $f$ , toujours en gardant le même esprit, mais en essayant d'obtenir une fonction Lipschitz. Ainsi définissons-nous

$$V(Q_{i,j}) := \sum_k \frac{D(X_k, (Q_{i,j})^c) - D(Y_k, (Q_{i,j})^c)}{D(Q_{i,j})} \quad (1.4)$$

, où  $D(x, A)$  est la distance entre le point  $x$  et l'ensemble  $A$  et  $D(A)$  est le diamètre de  $A$ , et

$$f(u) := \sum_{i,j} \frac{V(Q_{i,j})}{D(Q_{i,j})} D(u, (Q_{i,j})^c). \quad (1.5)$$

La fonction  $f$  ainsi obtenue est bien Lipschitz, mais de constante de Lipschitz trop grande (plus grande que  $(n \log n)^{\frac{1}{2}}$ ). Le but est maintenant est diminuer cette constante de Lipschitz pour arriver à  $(n \log n)^{\frac{1}{2}}$ . Pour ce faire nous allons redéfinir  $f$  une dernière fois, en faisant appel à des temps d'arrêt appropriés.

Pour  $u \in Q$ , nous notons  $Q_i^u$  le carré  $Q_{i,j}$  qui contient  $u$  et nous définissons

$$f_i(u) := \frac{V(Q_i^u)}{D(Q_i^u)} D(u, (Q_i^u)^c). \quad (1.6)$$

Puis, soit  $t(u)$  le temps d'arrêt défini comme étant le plus grand entier  $t \leq m$  tel que

$$\sup_{v \in Q_i^u} \max_{1 \leq k \leq t} \left| \sum_{i=1}^k f'_i(v) \right| \leq (Cn \log n)^{\frac{1}{2}}. \quad (1.7)$$

Enfin, la version finale de la fonction  $f$  est alors

$$f(u) := \sum_{i=1}^{t(u)} f_i(u). \quad (1.8)$$

La fonction  $f$  ainsi obtenue est continue, et différentiable par morceaux, et donc par (1.7), nous avons bien que  $f$  est Lipschitz, de constante de Lipschitz  $L = \sqrt{Cn \log n}$ .

Pour conclure, il nous reste donc à montrer que pour la fonction  $f$  ainsi définie, nous avons bien  $A(f) \geq Cn \log n$  avec probabilité  $1 - o(1)$ .

Pour cela, soit  $U_{i,j}$  égal à 1 si  $t(u) \geq i$  sur  $Q_{i,j}$  et 0 sinon. Il n'est pas difficile de voir qu'alors  $A(f) = \sum_{i,j} U_{i,j} V^2(Q_{i,j})$ . Par ailleurs, remarquons que la suite  $U_i := 4^{-i} \sum_j U_{i,j}$  est décroissante.

Enfin, il est possible de montrer, notamment à l'aide d'un résultat d'approximation par des ponts Browniens, et à des estimées sur des processus gaussiens, que  $U_m \geq \frac{1}{2}$ , et qu'avec probabilité  $1 - o(1)$ , nous avons  $\sum_{j=1}^{4^i} \epsilon_j V^2(Q_{i,j}) \geq Cn$  pour tout choix de  $\epsilon_j = 0, 1$  tels que  $\sum_{j=1}^{4^i} \epsilon_j \geq \frac{1}{2} 4^i$ .

En regroupant toutes les informations ci-dessus, nous pouvons alors conclure et aboutir au résultat annoncé. □

Pour la borne supérieure, nous référons à l'article original [AKT84].

## 1.2 Extension du résultat pour toute loi à support compact

Le théorème AKT est, comme nous l'avons vu en introduction puis dans la section précédente, un résultat concernant les lois uniformes sur le carré. Nous pouvons alors nous demander, de manière assez naturelle, si ce résultat s'étend pour d'autres lois, dans un premier temps

pour des lois à support compact. C'est précisément ce à quoi répond M. Talagrand dans [Tal92], en étendant le résultat du théorème AKT à toutes les lois à support dans le carré unité. Plus précisément, M. Talagrand démontre le résultat suivant.

**Théorème 1.2** (Théorème AKT pour toute loi à support compact). *Soit  $\mu$  une probabilité sur  $[0, 1]^2$ .*

*Soient  $X_i, Y_i$  iid de loi  $\mu$ .*

*Soit  $T_n = T_n(1) := \min_{\pi} \sum_{i=0}^n |X_i - Y_{\pi(i)}|$  où l'infimum est pris sur l'ensemble des permutations  $\pi$  de  $\{1, \dots, n\}$ .*

*Alors :*

$$\mathbb{E}\left(\frac{T_n}{n}\right) \leq C\left(\frac{\log n}{n}\right)^{\frac{1}{2}},$$

où  $C > 0$  est une constante universelle.

A partir de maintenant nous reprenons les notations "actuelles", qui sont en tout cas celles présentes dans les articles que nous avons le plus étudiés. Nous nous intéressons donc à  $\mathbb{E}(W_p^p(\mu_n, \nu_n))$  pour  $1 \leq p < +\infty$ . Comme souvent, certaines valeurs de  $p$  sont plus intéressantes, ou du moins plus faciles à étudier, que d'autres. En particulier, les cas  $p = 1$  et  $p = 2$  ont suscité l'intérêt de manière importante. Le cas  $p = 2$  est l'objet de la section suivante, dans laquelle nous présentons ce qui est sans doute le résultat phare depuis le théorème original de 1984, avec l'existence et l'exhibition d'une constante optimale en dimension 1 et 2. Nous ne donnons pas la preuve du théorème (1.2), notons toutefois qu'elle fait appel à une théorie plus générale développée par M. Talagrand, exposée notamment dans [Tal14], la théorie du "generic chaining", et qui a des applications diverses et plus larges que le seul théorème AKT.

Cependant, avant de passer à la section suivante, restons un peu sur le cas des lois générales à support compact, et revenons en particulier sur le cas  $p = 1$ . Nous présentons une approche différente, bien plus récente que celle de M. Talagrand, de traiter le problème dans ce cas spécifique. Cette approche est celle de S. Bobkov et M. Ledoux dans [BL19b], l'article qui était d'ailleurs à la base le sujet du stage, et fait appel à de l'analyse de Fourier. Nous nous sommes inspirés de cette approche et avons essayé de l'étendre, comme nous allons le voir dans la section (2.1). La preuve, au moins pour la borne supérieure, de S. Bobkov et M. Ledoux est particulièrement simple et vaut la peine d'être quelque peu détaillée, c'est pourquoi nous donnons ici les grandes lignes de la démonstration. Les auteurs travaillent sur le tore de dimension  $d$ ,  $\mathbb{T}^d$ , et vont se servir à la fois de la dualité de Kantorovich (théorème (4.2)) et de la décomposition en série de Fourier des fonctions périodiques afin d'arriver à une borne supérieure.

Enonçons tout d'abord le résultat de S. Bobkov et M. Ledoux, qui correspond au théorème 3 dans [BL19b].

**Théorème 1.3.** *Soient  $X_1, \dots, X_n, \dots, Y_1, \dots, Y_n, \dots$  des variables aléatoires iid à valeurs dans  $[0, 1]^d$ . Soit  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  et  $\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$ . Alors*

$$\mathbb{E}\left[W_1(\mu_n, \nu_n)\right] = \begin{cases} O\left(\frac{1}{\sqrt{n}}\right) & \text{si } d = 1, \\ O\left(\sqrt{\frac{\log n}{n}}\right) & \text{si } d = 2, \\ O\left(\frac{1}{n^{1/d}}\right) & \text{si } d \geq 3. \end{cases}$$

*Idée de preuve.* Comme annoncé, la preuve fait appel à de l'analyse de Fourier.

#### . Transformée de Fourier et Wasserstein

Si  $\mu$  est une mesure sur  $\mathcal{Q}^d := [-\pi, \pi]^d$ , nous définissons sa transformée de Fourier  $f_\mu$  comme la suite multi-indices pour  $m \in \mathbb{Z}^d$

$$f_\mu(m) := \int_{\mathcal{Q}^d} e^{im \cdot x} d\mu(x). \quad (1.9)$$

Si  $u$  est une fonction  $2\pi$ -périodique sur  $\mathbb{R}^d$  suffisamment régulière, par décomposition en série de Fourier, nous pouvons écrire

$$u(x) = \sum_{m \in \mathbb{Z}^d} a_m e^{im \cdot x}. \quad (1.10)$$

Ainsi, en différenciant par rapport à une coordonnée, puis en utilisant l'égalité de Parseval, et enfin en sommant sur les coordonnées, obtenons-nous

$$\frac{1}{(2\pi)^d} \int_{\mathcal{Q}^d} |\nabla u(x)|^2 dx = \sum_{m \in \mathbb{Z}^d} |m|^2 |a_m|^2. \quad (1.11)$$

Soient maintenant  $\mu$  et  $\nu$  deux probabilités sur  $\mathcal{Q}^d$ . Nous avons

$$\int_{\mathcal{Q}^d} u d\mu - \int_{\mathcal{Q}^d} u d\nu = \sum_{m \in \mathbb{Z}^d} a_m [f_\mu(m) - f_\nu(m)]. \quad (1.12)$$

Si de plus  $u$  est Lipschitz de constante de Lipschitz inférieure ou égale à 1, le terme de gauche dans (1.11) est inférieur ou égal à 1 (et donc celui de droite aussi), et donc en combinant cela avec (1.12) et en utilisant l'inégalité de Cauchy-Schwarz, nous obtenons :

$$\left| \int_{\mathcal{Q}^d} u d\mu - \int_{\mathcal{Q}^d} u d\nu \right|^2 \leq \sum_{m \neq 0} \frac{1}{|m|^2} |f_\mu(m) - f_\nu(m)|^2. \quad (1.13)$$

Enfin, en prenant le sup sur les fonctions  $u$   $2\pi$ -périodiques suffisamment régulières et Lipschitz de norme 1, et en utilisant la dualité de Kantorovich (4.2), nous obtenons :

$$W_1(\mu, \nu)^2 \leq \sum_{m \neq 0} \frac{1}{|m|^2} |f_\mu(m) - f_\nu(m)|^2. \quad (1.14)$$

#### . Régularisation et coût

Nous allons régulariser à l'aide d'un outil bien connu en analyse et en probabilités : la convolution. Nous convolons avec le noyau gaussien de la chaleur sur  $\mathcal{Q}^d$ . Plus précisément, sur  $\mathcal{Q}^d$ , nous considérons le noyau de la chaleur

$$p_t(x) := \frac{1}{(2\pi)^d} \sum_{m \in \mathbb{Z}^d} e^{im \cdot x - |m|^2 t}. \quad (1.15)$$

$p_t$  est la densité de la mesure de probabilité  $\gamma_t$ , à support dans  $\mathcal{Q}^d$ , dont la transformée de Fourier est donnée par  $f_{\gamma_t}(m) = e^{-|m|^2 t}$ . Si  $\mu$  est une probabilité sur  $\mathcal{Q}^d$ , nous définissons  $\mu_t$  la convolution de  $\mu$  avec  $\gamma_t$ , caractérisée par l'égalité

$$\int_{\mathcal{Q}^d} g d\mu_t = \int_{\mathcal{Q}^d} \int_{\mathcal{Q}^d} g(x+y) p_t(y) dy d\mu(x).$$

Remarquons tout de suite que

$$f_{\mu_t}(m) = e^{-|m|^2 t} f_\mu(m) \quad (1.16)$$

Nous devons contrôler le coût de cette régularisation en terme de distance de Wasserstein  $W_1$ . Pour cela, si  $u : \mathbb{R}^d \rightarrow \mathbb{R}$  est une fonction  $2\pi$ -périodique, 1-Lipschitz, nous avons

$$\left| \int_{\mathcal{Q}^d} u d\mu_t - \int_{\mathcal{Q}^d} u d\mu \right| \leq \int_{\mathcal{Q}^d} |y| p_t(y) dy. \quad (1.17)$$

Puis, en prenant le supremum sur les fonctions  $u$  1-Lipschitz, et en utilisant l'inégalité de Jensen, nous obtenons

$$W_1(\mu, \mu_t) \leq \left( \int_{\mathcal{Q}^d} |y|^2 p_t(y) dy \right)^{1/2} \leq 2dt. \quad (1.18)$$

Ensuite, en utilisant l'inégalité triangulaire pour  $W_1$ , en appliquant l'inégalité (1.14) à  $\mu_t$  et  $\nu_t$ , et l'inégalité (1.18), nous obtenons que pour toutes mesures  $\mu$  et  $\nu$  sur  $\mathcal{Q}^d$  et pour tout  $t > 0$ ,

$$W_1(\mu, \nu) \leq \left( \sum_{m \neq 0} \frac{e^{-2|m|^2 t}}{|m|^2} |f_\mu(m) - f_\nu(m)|^2 \right)^{\frac{1}{2}} + 2\sqrt{2dt}. \quad (1.19)$$

. **Application au théorème AKT** Soient  $X_1, \dots, X_n, Y_1, \dots, Y_n$  des variables aléatoires à valeurs dans  $[0, 1]^d$  telles que les couples  $(X_1, Y_1), \dots, (X_n, Y_n)$  sont deux à deux indépendants et telles que pour tout  $k$  les variables  $X_k$  et  $Y_k$  sont de même loi. En prenant l'espérance et en utilisant l'inégalité de Jensen dans (1.19), nous obtenons immédiatement

$$\mathbb{E}[W_1(\mu_n, \nu_n)] \leq \left( \sum_{m \neq 0} \frac{e^{-2|m|^2 t}}{|m|^2} \mathbb{E}[|f_{\mu_n}(m) - f_{\nu_n}(m)|^2] \right)^{\frac{1}{2}} + 2\sqrt{2dt}. \quad (1.20)$$

Or par les hypothèses sur les  $X_i$  et les  $Y_i$ , nous avons  $\mathbb{E}[|f_{\mu_n}(m) - f_{\nu_n}(m)|^2] \leq \frac{4}{n}$ .

Ce qui nous donne finalement

$$\mathbb{E}[W_1(\mu_n, \nu_n)] \leq \frac{2}{\sqrt{n}} \left( \sum_{m \neq 0} \frac{e^{-2|m|^2 t}}{|m|^2} \right)^{\frac{1}{2}} + 2\sqrt{2dt}. \quad (1.21)$$

En effectuant une comparaison série-intégrale pour estimer la série dans le terme de droite, et en optimisant ensuite en  $t > 0$  petit, nous obtenons bien le théorème (1.3).  $\square$

Nous ne donnons pas l'idée détaillée de la preuve de la borne inférieure, que les auteurs donnent aussi dans [BL19b], car celle-ci est assez technique, mais nous donnons tout de même le résultat principal sur lequel s'appuient les auteurs, que nous retrouvons aussi dans [AST19], qui est un théorème d'approximation des fonctions Sobolev par des fonctions Lipschitz, et qui est intéressant en tant que tel.

**Théorème 1.4** (Approximation de Lusin des fonctions Sobolev). *Pour tout  $p > 1$  et  $h \in H^{1,p}(\mathbb{T}^d)$ , et  $\alpha > 0$ , il existe une fonction  $\phi : \mathbb{T}^d \rightarrow \mathbb{R}$   $\alpha$ -Lipschitz telle que*

$$|\{h \neq \phi\}| \leq \frac{C(d, p)}{\lambda^p} \int_{\mathbb{T}^d} |\nabla h|^p dx.$$

La stratégie des auteurs est d'utiliser cette approximation pour une fonction  $h$  bien choisie, telle que  $\Delta h$  soit égal à la densité de  $\mu_n^t - \nu_n^t$  (ce qui fait le lien avec la section suivante), et ensuite utiliser la dualité de Kantorovich à la fonction  $\phi$  ainsi obtenue.

### 1.3 Constante optimale en dimension 1 et 2 : cas uniforme

Nous présentons ici un résultat majeur, démontré en 2019 par L. Ambrosio, F. Stra et D. Trevisan dans [AST19]. Le théorème concerne les lois uniformes sur le carré en dimension 2, et est plus précis que le théorème (1.1) car il donne la constante optimale.

**Théorème 1.5** (Constante optimale dans le cas  $p=2$ ). *Si  $D = [0, 1]^2$  ou si  $D$  est une variété Riemannienne, sans bord, de dimension 2. Soit  $\mu$  la mesure de volume sur  $D$ , et soit  $\mu_D := \frac{\mu}{\mu(D)}$ .*

*Soient  $\mu_n := \frac{1}{n} \sum_{i=0}^n \delta_{X_i}$  et  $\nu_n := \frac{1}{n} \sum_{i=0}^n \delta_{Y_i}$  où les  $X_i, Y_i$  sont iid de loi  $\mu_D$ . Alors :*

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E}(W_2^2(\mu_n, \mu_D)) = \frac{\mu(D)}{4\pi}. \quad (1.22)$$

*De plus, si  $D = [0, 1]^2$  ou  $D = \mathbb{T}^2$ , alors :*

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E}(W_2^2(\mu_n, \nu_n)) = \frac{1}{2\pi}. \quad (1.23)$$

Ce résultat a été une avancée significative, puisqu'elle a notamment permis de confirmer une des conjectures faites par S. Caracciolo, C. Lucibello, G. Parisi et G. Sicuro dans [CLPS14]. C'est aussi une avancée en terme de méthode utilisée, puisque les auteurs ont été les premiers à penser à régulariser par des noyaux de la chaleur et à se servir de la résolution d'une EDP comme étape de la preuve. Dans [AG19], L. Ambrosio et F. Glaudo parviennent à généraliser la deuxième partie du théorème, en se débarrassant de l'hypothèse  $D = [0, 1]^2$  ou  $D = \mathbb{T}^2$ . Ils arrivent ainsi à montrer que le résultat énoncé au dessus est en fait vrai pour toute variété compacte de dimension 2 et de volume égal à 1.

Nous allons donner ci-dessous une idée de la preuve de L. Ambrosio et F. Glaudo, que nous trouvons particulièrement intéressante, puisqu'elle fait appel à la résolution d'une EDP, ainsi qu'à des estimées fines sur les noyaux de la chaleur. Cependant, avant de passer à la preuve du résultat, revenons rapidement sur l'approche de S. Caracciolo, C. Lucibello, G. Parisi et G. Sicuro dans [CLPS14]. Cet ansatz est très simple et intuitif et permet d'effectuer des conjectures puissantes. Elle se fonde tout d'abord sur la formulation originale (4.2) du problème du transport optimal par Gaspard Monge. Nous nous plaçons sur le tore plat de dimension  $d$ ,  $\mathbb{T}^d$ , nous considérons deux mesures  $\mu_1$  et  $\mu_2$  admettant les densités respectives  $\rho_1$  et  $\rho_2$  et nous notons donc  $\mathcal{M}$  l'ensemble des fonctions  $T : \mathbb{T}^d \rightarrow \mathbb{T}^d$  telles que  $T\#\mu_1 = \mu_2$ .  $\mathcal{M}$  est de manière équivalente constitué des fonctions  $T$  telles que

$$\rho_2(x) = \rho_1(T(x)) \det J_T(x) \quad (1.24)$$

où  $J_T$  est la matrice jacobienne de  $T$ .

Dans le cas où  $c(x, y) = \|x - y\|^2$  est le coût quadratique, un résultat connu de transport optimal affirme que la fonction  $T$  optimale peut s'écrire comme le gradient d'une certaine fonction  $\phi$ , i.e  $T(x) = \nabla\phi(x)$ . Dans ce cas, l'équation (1.24) peut se réécrire de la façon suivante

$$\rho_2(x) = \rho_1(T(x)) \det \text{Hess } \phi(x) \quad (1.25)$$

où  $\text{Hess } \phi(x)$  est la matrice hessienne de  $\phi$ .

Supposons de plus que  $\mu_1$  et  $\mu_2$  soient des déformations de la loi uniforme, c'est-à-dire que  $\rho_i(x) = 1 + \delta f_i(x)$  avec  $|\delta f_i(x)| \ll 1$  pour  $i = 1, 2$ . Nous nous attendons à ce que dans ce cas, la fonction de transport optimale  $T$  puisse s'écrire  $T(x) = x + t(x)$ . Une approximation au premier ordre donne alors  $\det J_T(x) \approx 1 + \text{div } t(x)$ . L'équation (1.24) devient alors

$$\text{div } t(x) = \rho_1(x) - \rho_2(x) =: \delta f(x) \quad (1.26)$$

En particulier dans le cas quadratique, nous avons  $t(x) = \nabla\psi(x)$  pour  $\psi$  une certaine fonction et à partir de (1.26) nous obtenons la formule de Poisson

$$\Delta\psi = \delta f. \quad (1.27)$$

En notant  $\hat{f}_n = \int_{\mathbb{T}^d} \delta f(x) e^{-2i\pi n \cdot x} dx$ , pour  $n \in \mathbb{Z}^d$ , les coefficients de Fourier de  $f$ , nous obtenons que le coût du transport entre  $\mu_1$  et  $\mu_2$  est donné par

$$\int_{\mathbb{T}^d} |\nabla\psi(x)|^2 dx = \sum_{n \in \mathbb{Z}^d \setminus \{0\}} \frac{|\delta \hat{f}_n|^2}{4\pi^2 |n|^2}. \quad (1.28)$$

L'idée va alors de transposer l'équation (1.28) au cas des mesures empiriques. Bien que cela ne soit a priori pas rigoureux mathématiquement, cela donne des ordres de grandeur précis. Notons donc

$$\delta \hat{f}_n := \frac{1}{N} \sum_{i=1}^N (e^{2i\pi n \cdot X_i} - e^{2i\pi n \cdot Y_i}) \quad (1.29)$$

, pour  $n \in \mathbb{Z}^d$ , les coefficients de Fourier de la différence entre les deux mesures empiriques  $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{X_i}$  et  $\nu_N = \frac{1}{N} \sum_{i=1}^N \delta_{Y_i}$ . Un calcul direct donne  $\mathbb{E}[|\delta \hat{f}_n|^2] = \frac{2}{N}$ . En utilisant l'équation (1.28) et en prenant l'espérance, nous en déduisons que

$$N^{\frac{2}{d}} \mathbb{E}[W_2^2(\mu_N, \nu_N)] \sim \sum_{n \in \mathbb{Z}^d \setminus \{0\}} \frac{N^{\frac{2-d}{d}}}{2\pi^2 |n|^2}. \quad (1.30)$$

Malheureusement, cette somme ne converge que pour  $d = 1$ , et dans ce cas cette approche donne la bonne constante optimale. Pour le cas  $d \geq 2$ , les auteurs procèdent à une régularisation de cette somme, et qui finit par leur donner le bon ordre en toute dimension, et même à conjecturer la constante optimale et la vitesse de convergence en dimension 2.

Passons maintenant à l'idée de la preuve du théorème (1.5), qui confirme la conjecture des physiciens.

*Idée de preuve du théorème (1.5).*

### . Régularisation

Comme sous-entendu précédemment, il est plus simple de considérer des distances de Wasserstein entre des objets lisses, plutôt qu'entre des mesures empiriques, qui sont des objets dégénérés, c'est pourquoi nous régularisons nos mesures empiriques, à l'aide de noyaux de la chaleur. Pour un temps  $t > 0$ , nous notons donc

$$\mu_n^t := P_t^* \mu_n = \left( \frac{1}{n} \sum_{i=1}^n p_t(X_i, \cdot) \right) \mu = u_n^t \mu \quad (1.31)$$

, où  $u_n^t$  est définie comme la densité de  $\mu_n^t$  par rapport à  $\mu$ , et où  $p_t(\cdot, \cdot)$  est le noyau de la chaleur sur  $M$  au temps  $t$ .

Nous définissons maintenant le potentiel  $f_n^t : M \rightarrow \mathbb{R}$ , comme l'unique fonction de moyenne nulle telle que

$$\Delta f_n^t = u_n^t - 1 \quad (1.32)$$

(où  $\Delta$  est l'opérateur de Laplace-Beltrami sur  $M$ ).

Afin de simplifier l'expression de certains termes dans les inégalités à venir, nous décidons de nous placer sur un certain événement, dont nous montrons qu'il est très probable.

Plus précisément, soit

$$A_\xi^{n,t} := \{ \|\nabla^2 f_n^t\|_\infty < \xi \} \quad (1.33)$$

En particulier, par (1.32) et comme  $\|\Delta f\|_\infty \leq C \|\nabla^2 f\|_\infty$ , nous avons que l'évènement  $A_\xi^{n,t}$  est contenu dans  $\{ \|u_n^t - 1\|_\infty < \xi \}$ .

Nous montrons alors qu'il existe  $a = a(M) > 1$  tel que pour tout  $n \in \mathbb{N}$ , pour tout  $0 < \xi < 1$ , et pour tout  $0 < t < 1$ , l'équation suivante soit vraie :

$$\mathbb{P}((A_\xi^{n,t})^c) \leq \frac{C}{\xi^2 t^3} a^{-nt\xi^2}. \quad (1.34)$$

Il s'agira alors d'effectuer une bonne optimisation en  $t$  et  $\xi$  afin de pouvoir vraiment négliger le complémentaire de l'évènement  $A_\xi^{n,t}$ .

Nous remarquons par ailleurs que c'est le seul moment de la preuve où interviennent des arguments de nature réellement probabilistes avec des inégalités classiques de concentration, dont l'inégalité de Bernstein, ce qui est une preuve de la nature plus analytique de cette preuve, comparée à la preuve originale du théorème. Pour plus d'informations sur les différentes inégalités de concentration, nous référons par exemple à [BLM16].

Ensuite, un outil fondamental et extrêmement utile de la preuve est la formule de la trace pour les noyaux de la chaleur (que nous retrouverons aussi dans d'autres sections ensuite) qui s'exprime de la manière suivante :

$$\int_M (p_t(x, x) - 1) d\mu = \frac{1}{4\pi t} + O\left(\frac{1}{\sqrt{t}}\right) \quad (1.35)$$

Cette formule de la trace nous notamment d'obtenir des estimées sur le noyau de la chaleur et ses dérivées successives.

Après plusieurs étapes de calcul, nous arrivons à l'approximation suivante, qui sera essentielle dans la suite de la preuve :

$$\mathbb{E} \left[ \int_M |\nabla f_n^t|^2 d\mu \right] = \frac{|\log t|}{4\pi n} + O\left(\frac{1}{n}\right). \quad (1.36)$$

### . Coût du transport

Le but est ensuite de majorer la distance de Wasserstein  $W_2$  entre  $\mu_n^t$  et  $\mu$  en se servant d'une formule célèbre en transport optimal : la formule de Benamou-Brenier (théorème (4.3) en appendice).

Avec un peu de travail à partir de cette formule, nous en déduisons les deux inégalités suivantes : étant données deux densités  $u_0, u_1$ , si  $f \in C^\infty(M)$  est l'unique fonction de moyenne nulle telle que  $-\Delta f = u_1 - u_0$ , alors :

$$W_2^2(u_0\mu, u_1\mu) \leq 4 \int_M \frac{|\nabla f|^2}{u_0} d\mu, \quad (1.37)$$

et aussi :

$$W_2^2(u_0\mu, u_1\mu) \leq 4 \int_M |\nabla f|^2 \frac{\log u_1 - \log u_0}{u_1 - u_0} d\mu. \quad (1.38)$$

C'est pour l'utilisation de ces inégalités qu'il sera judicieux de se placer sur l'évènement  $A_\xi^{n,t}$ . Nous voyons tout de suite le lien avec l'étape précédente de régularisation : en se plaçant donc sur l'évènement  $A_\xi^{n,t}$ , si nous regardons la distance de Wasserstein  $W_2^2$  entre  $\mu_n^t$  et  $\mu$ , nous aurons une inégalité du type  $W_2^2(\mu_n^t, \mu) \leq C \int_M |\nabla f_n^t|^2 d\mu$ , où  $f_n^t$  est le potentiel défini en (1.32). Il s'agit ensuite d'utiliser l'équation (1.36) et enfin d'optimiser astucieusement en  $t > 0$  petit. Cette optimisation est d'ailleurs l'objet de l'étape suivante, où nous raffinons une inégalité classique, afin de pouvoir optimiser différemment.

### . Contractivité raffinée du noyau de la chaleur

L'étape suivante est de majorer la distance de Wasserstein  $W_2$  entre  $\mu_n$  et  $\mu_n^t = P_t^* \mu_n$ .

Une inégalité assez classique, et que nous retrouvons notamment dans [AST19] ou encore dans [Led19a], est :

$$W_2^2(P_t^* \mu, \mu) \leq Ct \quad (1.39)$$

où  $C$  est une constante et  $t > 0$ . Notre but est alors de raffiner cette inégalité, afin de pouvoir optimiser d'une meilleure façon en  $t$  petit.

Nous arrivons alors au résultat suivant : il existe une constante  $C = C(M) > 0$  telle que si  $t = \frac{\alpha}{n}$ , avec  $\alpha \geq C \log n$ , alors

$$\mathbb{E} [W_2^2(\mu_n^t, \mu_n)] \leq C' \frac{\log \alpha}{n} \quad (\ll t = \frac{\alpha}{n}). \quad (1.40)$$

La preuve de ce résultat est fondée sur plusieurs utilisations successives de l'inégalité (1.39).

Comme annoncé à la fin de l'étape précédente, l'inégalité (1.40) nous permet de procéder à une optimisation en  $t > 0$  de l'ordre de  $\frac{\log^3 n}{n}$ , alors que si nous nous étions contenté de l'inégalité (1.39), nous aurions dû prendre  $t = o\left(\frac{\log n}{n}\right)$  ce qui aurait causé des complications et n'aurait pas permis de conclure.

Avec toutes les inégalités présentées dans les trois premières étapes de la preuve, nous sommes désormais prêts à passer à la conclusion du théorème.

• **Conclusion**

Nous donnons ici seulement une idée de la preuve dans le cas du matching "unipartite"; la preuve dans le cas du matching bipartite étant très similaire, seulement un peu plus technique.

La stratégie est d'utiliser l'équation (1.36), avec un  $t = t(n)$  bien choisi (de l'ordre de  $\frac{\log^3 n}{n}$ ), pour approximer  $\frac{1}{4\pi}$  par  $\int_M |\nabla f_n^t|^2 d\mu$ , puis d'utiliser un résultat de transport optimal, dont vous pouvez trouver la preuve par exemple dans [Gla19] (corollaire 3.4), qui affirme que :

$$\int_M |\nabla f_n^t|^2 d\mu = W_2^2(\exp(\nabla f_n^t)\#\mu, \mu). \quad (1.41)$$

Ce résultat permet d'utiliser des inégalités de type triangulaire pour  $W_2$ , et alors, en réussissant à garder sous contrôle  $W_2^2(\exp(\nabla f_n^t)\#\mu, \mu_n^t)$ , nous obtenons bien le résultat annoncé à savoir

$$\lim_{n \rightarrow \infty} \frac{n}{\log n} \mathbb{E}(W_2^2(\mu_n, \mu_D)) = \frac{\mu(D)}{4\pi}. \quad (1.42)$$

□

## 1.4 Constante optimale en dimension 3 et plus : cas uniforme

A la fin de leur article, L. Ambrosio, D. Trevisan et F. Stra font remarquer que leur démarche échoue en dimension supérieure ou égale à 3, et ne permet ni à donner la limite, ni même à prouver son existence. Il paraît pourtant naturel de penser que cette limite existe bien, et c'est ce qu'affirment M. Goldman et D. Trevisan dans [GT20]. Les deux auteurs parviennent à montrer l'existence de la limite dans le cas uniforme sur le cube en dimension 3 et plus, et ce pour tout  $p \geq 1$ . Plus précisément ils démontrent le théorème suivant :

**Théorème 1.6** (Existence de la constante optimale en dimension  $d \geq 3$ ). *Soit  $d \geq 3$ , et  $p \geq 1$ .*

*Soient  $X_i, Y_i$  iid de loi uniforme sur  $[0, 1]^d$ , et soient  $\mu_n = \frac{1}{n} \sum_{i=0}^n \delta_{X_i}$  et  $\nu_n = \frac{1}{n} \sum_{i=0}^n \delta_{Y_i}$ .*

*Il existe une constante  $C_\infty = C_\infty(p, d)$  telle que :*

$$\lim_{n \rightarrow \infty} n^{\frac{p}{d}} \mathbb{E}\left(W_p^p(\mu_n, \nu_n)\right) = C_\infty.$$

Bien que très récente, la preuve donnée par M. Goldman et D. Trevisan paraît de nature moins analytique que l'article évoqué dans la section précédente. En effet, même si nous y retrouvons à un moment l'idée d'utiliser une EDP pour avoir une inégalité en distance de Wasserstein, ce n'est pas la même utilisation que celle vue précédemment. Par ailleurs, les autres arguments utilisés paraissent être de nature plus probabiliste. Ainsi les auteurs utilisent-ils des arguments de sous-additivité, un processus de poissonisation, puis de dépoissonisation du problème, qui sont des outils assez classiques en probabilités. Là aussi nous ne donnons l'idée de la preuve que pour le matching de la mesure empirique à la mesure de référence, car les idées sont sensiblement les mêmes que dans le cas bipartite, mais les calculs sont moins techniques. Remarquons cependant que leur théorème échoue à donner la valeur de la constante optimale, ce qui est un défaut plus général des preuves faisant appel à des arguments de sous-additivité, qui de manière générale donnent l'existence d'une limite pouvant s'écrire sous la forme d'un infimum mais ne donnent pas l'expression de cette limite.

*Idée de preuve.*

• **Sous-additivité de Wasserstein**

Pour cette étape nous considérons  $\Omega$  un borélien quelconque de  $\mathbb{R}^d$ . Nous allons, tout au long de la preuve, avoir affaire avec des partitions, et donc des distances de Wasserstein sur différents domaines. Par souci de clarté, nous noterons donc  $W_\Omega^p$  la distance de Wasserstein  $W_p$  sur  $\Omega$ , élevée à la puissance  $p$ . Nous commençons en douceur, en rappelant que si  $\mu_i(\Omega) = \nu_i(\Omega)$  pour tout  $i$  alors  $W_p^p(\sum_i \mu_i, \sum_i \nu_i) \leq \sum_i W_p^p(\mu_i, \nu_i)$ .

En combinant cette inégalité avec l'inégalité triangulaire pour  $W_p$ , nous arrivons facilement à l'inégalité de sous-additivité suivante : pour tout  $p \geq 1$ , il existe une constante  $C = C(p) > 0$  telle que pour tout  $\Omega \in \mathcal{B}(\mathbb{R}^d)$ , pour toute partition  $(\Omega_i)_i$  de  $\Omega$ , pour toutes mesures  $\mu$  et  $\lambda$  sur  $\Omega$  et pour tout  $\epsilon \in (0, 1)$ , nous ayons

$$W_\Omega^p(\mu, \frac{\mu(\Omega)}{\lambda(\Omega)}\lambda) \leq (1 + \epsilon) \sum_i W_{\Omega_i}^p(\mu, \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\lambda) + \frac{c}{\epsilon^{p-1}} W_\Omega^p(\sum_i \frac{\mu(\Omega_i)}{\lambda(\Omega_i)}\lambda \mathbf{1}_{\Omega_i}, \frac{\mu(\Omega)}{\lambda(\Omega)}\lambda). \quad (1.43)$$

Cette inégalité, bien que lourde à première vue, est en fait très simple à obtenir, et sera utile pour montrer l'inégalité cruciale (1.45).

#### . Matching d'un processus de Poisson à la mesure de Lebesgue

L'idée de cette étape est de s'intéresser à l'espérance de la distance de Wasserstein entre un processus de Poisson d'intensité 1 restreint à un cube de côté  $L$  et la mesure de Lebesgue sur ce même cube. Plus précisément, nous notons  $Q_L = [0, L]^d$ , soit  $\mu$  un processus de Poisson d'intensité 1, nous définissons

$$f^{ref}(L) := \mathbb{E}\left[\frac{1}{|Q_L|} W_{Q_L}^p(\mu, \kappa)\right], \quad (1.44)$$

où  $\kappa = \frac{\mu(Q_L)}{|Q_L|}$ .

Le but de cette étape est de montrer que la limite de  $f^{ref}$  en  $+\infty$  existe.

Pour cela, nous montrons le résultat intermédiaire de sous-additivité suivant, et pour lequel nous nous servons de l'inégalité (1.43) : pour tout  $d \geq 3$  et tout  $p \geq 1$ , il existe une constante  $C > 0$  telle que pour tout  $L \geq 1$  et tout  $m \in \mathbb{N}$ , nous ayons

$$f^{ref}(mL) \leq f^{ref}(L) + \frac{C}{L^{\frac{d-2}{2}}}. \quad (1.45)$$

C'est ici que, pour la première fois depuis le début de la preuve, il est nécessaire de supposer que  $d \geq 3$  (comme le laisse supposer le terme en  $d - 2$ ), en effet pour aboutir à l'inégalité (1.45) il est essentiel de faire cette hypothèse.

Cette inégalité, couplée au fait que  $f^{ref}$  soit continue, permet d'utiliser un lemme de sous-additivité du type lemme de Fekete et d'aboutir au résultat suivant : pour tout  $d \geq 3$  et  $p \geq 1$ , la limite

$$f_\infty^{ref} := \lim_{L \rightarrow +\infty} f^{ref}(L) \quad (1.46)$$

existe et est strictement positive.

L'étape suivante est un processus de dépoissonisation, qui consiste donc à rendre à nouveau déterministe le nombre de points.

#### . Dépoissonisation

Nous montrons d'abord un lemme général de dépoissonisation, qui affirme que sous certaines hypothèses sur une fonction, la limite est la même quand on prend l'espérance de cette fonction appliquée à des variables de Poisson et quand on prend la limite classique en  $n \rightarrow +\infty$ . Plus précisément, soit  $f : (0, +\infty) \times \mathbb{N}^k \rightarrow [0, +\infty)$  vérifiant les trois conditions suivantes :

1.  $f(L|n) = L^p f(1|n)$  pour tout  $L > 0$  et  $n \in \mathbb{N}^k$  (p-homogénéité)
2.  $f(1|n) \leq C$  pour tout  $n \in \mathbb{N}^k$  (bornitude)
3.  $f(1|n) \leq f(1|m)$  pour tous  $m, n \in \mathbb{N}^k$  tels que  $m_i \leq n_i$  pour tout  $i = 1, \dots, k$  (monotonie).

Nous définissons  $f(L) := \mathbb{E}[f(L|N_L)]$  où  $N_L = (N_{L,i})_{i=1, \dots, k}$  où les  $N_{L,i}$  sont des variables aléatoires iid de loi de Poisson de paramètre  $L^d$ . Alors

$$\liminf_{n \rightarrow +\infty} n^{\frac{p}{d}} f(1|(n, \dots, n)) = \liminf_{L \rightarrow +\infty} f(L) \quad \text{et} \quad \limsup_{n \rightarrow +\infty} n^{\frac{p}{d}} f(1|(n, \dots, n)) = \limsup_{L \rightarrow +\infty} f(L). \quad (1.47)$$

Le but de la dernière étape va alors être d'appliquer ce procédé de dépoissonisation au problème de matching, ce qui permettra de conclure et d'arriver au théorème (1.6).

. **Conclusion**

Afin d'arriver au théorème (1.6), nous avons envie de combiner les résultats (1.46) et (1.47). Ainsi définissons-nous

$$f(L|n) := L^p \mathbb{E}[W_{Q_1}^p(\mu_n, 1)]$$

. Il n'est pas difficile de voir (même si la monotonie demande un peu de travail) que cette fonction vérifie les trois hypothèses requises pour notre résultat de dépoissonisation (1.47). Nous pouvons donc appliquer notre résultat de dépoissonisation; cependant celui-ci ne nous donne pas encore l'existence de la limite, c'est pour cette raison que nous le combinons avec l'équation (1.46). Pour conclure, il s'agit enfin de remarquer que

$$\lim_{L \rightarrow +\infty} \mathbb{E}[|f(L|N_L) - f^{ref}(L|N_L)|] = 0. \quad (1.48)$$

Ainsi, en combinant (1.46), (1.47) et (1.48), et en appliquant tout ça à notre fonction  $f$  nous retrouvons le résultat voulu, à savoir l'existence de la constante optimale en dimension 3 et plus.

□

## 1.5 Cas gaussien

Une fois les asymptotiques pour les lois à support compact comprises, il semble naturel de se demander ce qu'il en est pour des loi à support non compact. Bien évidemment, comme souvent en probabilités, la première loi à laquelle il est naturel de s'intéresser alors est la loi gaussienne standard sur  $\mathbb{R}^d$ .

Ce cas a notamment été étudié par M. Ledoux dans [Led19a] et [Led19b], puis par M. Ledoux et J.X. Zhu dans [LZ19]. Avant cela, M. Talagrand s'était aussi intéressé au cas gaussien en conjecturant notamment le théorème (1.7) dans [Tal18]. Le cas gaussien présente des cas particulièrement intéressants, par exemple nous observons une rupture pour le cas  $p = d$ ; rupture que nous ne retrouvons pas dans le cas uniforme. Nous donnons ici les résultats connus à ce jour pour les lois gaussiennes, et nous y revenons dans la section (2.3), en essayant de raffiner certains des résultats de cette section, à la recherche d'une éventuelle constante optimale.

En combinant les résultats de [Led19a] et [Led19b], qui s'intéressent à la distance de Wasserstein  $W_2$  entre une mesure empirique  $\mu_n$  et la mesure de référence  $\mu$  dans le cas de la dimension 2, le résultat peut s'énoncer de la manière suivante.

**Théorème 1.7.** *Soit  $\mu$  une loi gaussienne standard sur  $\mathbb{R}^2$ , et soit  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ , où les  $X_i$  sont iid de loi  $\mu$ .*

*Alors*

$$\mathbb{E}(W_2^2(\mu_n, \mu)) \approx \frac{(\log n)^2}{n}.$$

Nous observons tout de suite une différence avec le cas uniforme puisque nous avons un facteur  $\log n$  supplémentaire qui apparaît. En fait ce facteur supplémentaire semble être typique du cas où  $p = d$  qui est particulier dans le cas gaussien, et donc particulièrement intéressant. En effet, dans [LZ19], M. Ledoux et J. X. Zhu montrent le résultat suivant.

**Théorème 1.8.** *Soit  $d \geq 3$ , soit  $\mu$  la loi gaussienne standard sur  $\mathbb{R}^d$  et  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  où les  $X_i$  sont iid de loi  $\mu$ .*

*Alors, pour tout  $1 \leq p < d$ ,*

$$\mathbb{E}(W_p^p(\mu_n, \mu)) \approx \frac{1}{n^{p/d}}.$$

Ainsi, quand la dimension est plus grande que 3 et que  $p < d$ , retrouvons-nous les ordres de grandeur du théorème AKT standard. Le cas  $p = d$  est quelque peu plus délicat, et pour l'instant il n'a pas encore été prouvées que les bornes connues et données dans [LZ19] sont optimales. Pour l'instant le résultat connu, démontré dans [LZ19] est le suivant.

**Théorème 1.9.** *Soit  $d \geq 2$ , soit  $\mu$  la loi gaussienne standard sur  $\mathbb{R}^d$  et  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  où les  $X_i$  sont iid de loi  $\mu$ .*

Alors

$$\mathbb{E}(W_d^d(\mu_n, \mu)) \leq C_d \frac{(\log n)^{\kappa_d}}{n}$$

où

$$\kappa_d = \begin{cases} \frac{d^2+6d}{8} & \text{si } d = 2, 3 \\ \frac{d^2}{2} - \frac{3d}{4} & \text{si } d \geq 4. \end{cases}$$

Dans le cas où  $p = d = 2$ , nous retrouvons ainsi la borne supérieure dans le théorème (1.7). Par ailleurs une conjecture possible serait d'affirmer que

$$\mathbb{E}(W_d^d(\mu_n, \mu)) \approx \frac{(\log n)^{\frac{d}{2}}}{n}$$

pour  $p = d \geq 3$ , mais celle-ci n'est à notre connaissance pas prouvée à l'heure actuelle.

Par ailleurs, à ce jour, aucune conjecture n'est faite pour le cas  $p > d \geq 2$ .

Nous donnons ici une brève idée de la preuve du théorème (1.7), et nous y revenons dans la section (2.3) dans laquelle le but est de rendre plus fin les estimées dans le cas gaussien afin d'obtenir une éventuelle constante optimale.

*Idee de preuve du théorème (1.7).* Une des idées fondamentales de l'article est encore une inégalité fondée sur une EDP réalisée par un potentiel.

Soit  $\mu$  une loi gaussienne standard sur  $\mathbb{R}^d$ , avec  $d \geq 1$ . En fait une partie des résultats que nous énonçons dans cette démonstration peuvent s'appliquer plus généralement à des mesures  $\mu$  de la forme  $d\mu = e^{-V} dx$  où  $dx$  est la mesure de volume sur la variété sur laquelle on travaille.

#### . Localisation

Comme nous travaillons avec des mesures à support non compacts, il va être pratique de se ramener au cas compact. Un point important de la preuve est donc la localisation : plus précisément, pour  $R > 0$ , soit  $d\mu^R = \frac{\mathbf{1}_{B_R}}{\mu(B_R)} d\mu$ , où  $B_R$  est la boule euclidienne de centre 0 et de rayon  $R$ .

Nous définissons des variables iid  $X_i^R$ ,  $i \geq 0$ , de loi  $\mu^R$ . Nous définissons ensuite la mesure empirique localisée

$$\mu_n^R := \frac{1}{n} \sum_{i=1}^n \delta_{X_i^R}. \quad (1.49)$$

Nous avons alors,

$$\mathbb{E}[W_2^2(\mu_n, \mu_n^R)] \leq 4 \int_{|x|>R} |x|^2 d\mu \approx_{R \rightarrow \infty} R^d e^{-R^2/2}.$$

Un choix naturel d'optimisation pour  $R > 0$  est alors  $R = R(n) = c\sqrt{\log n}$  avec un  $c > 0$  suffisamment grand de telle sorte que

$$\mathbb{E}[W_2^2(\mu_n, \mu_n^R)] = O\left(\frac{1}{n^{c'}}\right) \quad (1.50)$$

pour un  $c' > 1$ .

• **Régularisation et coût de la régularisation**

Une des idées principales de la démonstration est encore la régularisation des mesures empiriques. Cependant, au lieu d'utiliser le noyau de la chaleur "standard" associé au laplacien, il est plus efficace pour des lois gaussiennes d'utiliser le noyau de Mehler, associé au semi-groupe d'Ornstein-Uhlenbeck, pour lequel la loi gaussienne standard est la mesure invariante. Nous référons par exemple à [BGL14] pour plus de précisions sur ce semi-groupe, que nous présentons aussi brièvement dans la section (4.3).

Nous notons donc  $p_t$  le noyau de Mehler, donné par

$$p_t(x, y) = \frac{1}{(1 - e^{-2t})^{d/2}} \exp\left(-\frac{e^{-2t}}{2(1 - e^{-2t})} [|x|^2 + |y|^2 - 2e^t x \cdot y]\right).$$

Vient maintenant l'étape de régularisation, pour  $t > 0$ , nous définissons

$$f_n^{R,t}(y) = \frac{1}{n} \sum_{i=1}^n p_t(X_i^R, y)$$

puis

$$d\mu_n^{R,t} = f_n^{R,t} d\mu. \tag{1.51}$$

Assez facilement, nous arrivons à l'estimée suivante, qui est une estimée assez classique :

$$\mathbb{E}[W_2^2(\mu_n^R, \mu_n^{R,t})] \leq 4dt. \tag{1.52}$$

• **Coût du transport et conclusion**

Il reste enfin à évaluer  $\mathbb{E}[W_2^2(\mu_n^{R,t}, \mu)]$ . Pour cela les auteurs montrent en fait d'abord un résultat plus général, qui ressemble à ce qui se trouve dans les articles [AST19] et [AG19].

Plus précisément, ils montrent que si  $1 \leq p < +\infty$  et si  $dv = f d\mu$ , alors

$$W_p(v, \mu) \leq p \|f - 1\|_{H^{-1,p}(\mu)} \tag{1.53}$$

où  $\|g\|_{H^{-1,p}(\mu)}$  est donnée par :

$$\|g\|_{H^{-1,p}(\mu)} := \left( \int_M |\nabla((-L)^{-1}g)|^p \right)^{1/p}. \tag{1.54}$$

En particulier dans le cas  $p = 2$ , un calcul rapide donne :

$$\|g\|_{H^{-1,2}(\mu)}^2 = 2 \int_0^\infty \int_M (P_t g)^2 d\mu dt \tag{1.55}$$

En appliquant ça à  $v = \mu_n^{R,t}$  nous obtenons en calculant  $P_t f_n^{R,t}$  :

$$\mathbb{E}[W_2^2(\mu_n^{R,t}, \mu)] \leq 4 \int_0^\infty \int_M \mathbb{E}[(P_t(f_n^{R,t} - 1))^2] d\mu dt. \tag{1.56}$$

Enfin, après des calculs techniques en distinguant les cas  $t$  petit et  $t$  grand, puis après optimisation, nous trouvons le bon ordre en dimension 2.  $\square$

## 2 Quelques avancées personnelles

Nous présentons dans cette section les résultats des recherches personnelles effectuées au long de ce stage, ayant abouties à des résultats plus ou moins significatifs.

## 2.1 Borne supérieure pour des lois générales

Le but de cette partie est de démontrer le théorème suivant, qui donne une borne supérieure de l'espérance de la distance de Wasserstein  $W_1$  entre deux mesures empiriques, où toutes les variables aléatoires sont supposées indépendantes.

**Théorème 2.1** (Borne supérieure). *Soit  $\mu \in P(\mathbb{R}^d)$  admettant un moment d'ordre 2 et  $X$  une variable aléatoire de loi  $\mu$ .*

*Si  $\forall R > 0, \mathbb{P}(\|X\|_\infty > R) \leq \Phi(R)$  où  $\Phi$  est une fonction strictement décroissante, alors*

$$\mathbb{E} \left[ W_1(\mu_n, \nu_n) \right] \leq \begin{cases} C \frac{1}{n^{1/2}} \left( \Phi^{-1} \left( \frac{1}{n} \right) \right)^{1/2} & \text{si } d = 1, \\ C \left( \frac{\log n}{n} \right)^{1/2} \left( \Phi^{-1} \left( \frac{\log n}{n} \right) \right)^{1/2} & \text{si } d = 2, \\ C n^{-1/d} \left( \Phi^{-1} \left( n^{-2/d} \right) \right)^{1/2} & \text{si } d \geq 3. \end{cases}$$

Nous avons des termes de correction par rapport aux résultats connus dans le cas compact traité par exemple dans [Tal92].

Ces résultats appliqués aux lois gaussiennes standards aboutissent à des résultats plus faibles que ceux montrés par M. Ledoux dans [Led19a], [Led19b] et par M. Ledoux et J.X. Zhu [LZ19] mais permettent par contre d'obtenir une borne supérieure pour des lois plus générales que les gaussiennes.

L'approche utilisée ici est grandement inspirée de celle de S. Bobkov et M. Ledoux dans [BL19b]. Nous reprenons en effet la démarche utilisée dans le cas de l'hypercube  $[0, 1]^d$ , et nous l'étendons à l'espace tout entier en travaillant dans des cubes de plus en plus grand. Nous effectuons des troncatures de notre mesure de départ, et utilisons de l'analyse de Fourier sur les cubes de plus en plus grand. Le terme de correction présent dans le théorème (2.1) provient justement de cette analyse de Fourier, et plus précisément de l'égalité de Parseval. Un premier constant que nous pouvons faire est que le résultat obtenu fait intervenir la queue de distribution de la loi considérée. Plus la queue est légère, plus la borne obtenue sera proche de la distance dans le cas uniforme. Nous en déduisons immédiatement les corollaires suivants, lorsque la queue de distribution est sous-polynomiale ou sous-exponentielle.

**Corollaire 2.1** (Queue sous-polynomiale). *Soit  $d \geq 1$ .*

*Si  $\mathbb{P}(\|X\|_\infty > R) \leq CR^{-\alpha}$ , pour un  $\alpha > 0$ , alors :*

$$\mathbb{E} \left[ W_1(\mu_n, \nu_n) \right] \leq \begin{cases} C n^{-\frac{1}{2} + \frac{1}{2\alpha}} & \text{si } d = 1, \\ C \left( \frac{\log n}{n} \right)^{\frac{1}{2}} \left( \frac{n}{\log n} \right)^{\frac{1}{\alpha}} & \text{si } d = 2, \\ C n^{-\frac{1}{d} + \frac{1}{d\alpha}} & \text{si } d \geq 3. \end{cases} \quad (2.1)$$

*En particulier, si  $\mathbb{P}(\|X\|_\infty > R)$  décroît plus vite que tout polynôme alors pour tout  $\epsilon > 0$ , nous obtenons :*

$$\mathbb{E} \left[ W_1(\mu_n, \nu_n) \right] \leq \begin{cases} C n^{-\frac{1}{2} + \epsilon} & \text{si } d = 1, \\ C \left( \frac{\log n}{n} \right)^{\frac{1}{2}} \left( \frac{n}{\log n} \right)^\epsilon & \text{si } d = 2, \\ C n^{-\frac{1}{d} + \epsilon} & \text{si } d \geq 3. \end{cases} \quad (2.2)$$

**Corollaire 2.2** (Queue sous-exponentielle). *Soit  $d \geq 1$*

*Si  $\mathbb{P}(\|X\|_\infty > R) \leq ce^{-R^\alpha}$ , pour un  $\alpha > 0$ , alors :*

$$\mathbb{E} \left[ W_1(\mu_n, \nu_n) \right] \leq \begin{cases} Cn^{-\frac{1}{2}}(\log n)^{\frac{1}{2\alpha}} & \text{si } d = 1, \\ C\left(\frac{\log n}{n}\right)^{\frac{1}{2}} \left(\log\left(\frac{n}{\log n}\right)\right)^{\frac{1}{2\alpha}} & \text{si } d = 2, \\ Cn^{-\frac{1}{d}}\left(\frac{2}{d}\log n\right)^{\frac{1}{d\alpha}} & \text{si } d \geq 3. \end{cases} \quad (2.3)$$

Avant de passer à la preuve du théorème (2.1), donnons deux exemples de ce que donne le résultat pour des lois à support non compact dont la queue de distribution est explicite et dont l'inverse est particulièrement simple à calculer.

*Exemple.* Appliquons notre résultat dans le cas d'une loi à queue lourde, mais qui admet quand même un moment d'ordre 2, par exemple une loi de Pareto de paramètres 1 et  $k$  avec  $k > 2$ .

Rappelons que si  $X$  suit une loi de Pareto de paramètres 1 et  $k$ , alors  $X$  admet une densité donnée par  $f(x) = \frac{k}{x^{k+1}}$  pour  $x \geq 1$ .

Et donc  $\mathbb{P}(X > R) = \frac{1}{R^k}$ , puis  $\Phi^{-1}(1/n) = n^{1/k}$ .

Ainsi, en appliquant le théorème 1.1 en dimension 1, nous obtenons, en reprenant les notations pour les mesures empiriques,

$$\mathbb{E} \left[ W_1(\mu_n, \nu_n) \right] \leq C \frac{1}{n^{\frac{k-1}{2k}}}.$$

*Exemple.* Traitons un dernier exemple avant de passer à la section suivante : celui d'une loi de Weibull de paramètres  $\lambda > 0$  et  $k > 0$ .

Une densité d'une variable aléatoire de loi de Weibull de paramètres  $\lambda > 0$  et  $k > 0$  est donnée par  $f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$  et alors un calcul direct donne  $\Phi(R) = e^{-\left(\frac{R}{\lambda}\right)^k}$ .

Ainsi  $\Phi^{-1}(1/n) = \lambda(\log n)^{1/k}$  et donc en appliquant le théorème 1.1 en dimension 1 nous obtenons :

$$\mathbb{E} \left[ W_1(\mu_n, \nu_n) \right] \leq C \frac{\sqrt{\lambda}(\log n)^{1/2k}}{n^{1/2}}.$$

Nous donnons maintenant une idée de la preuve du théorème (2.1).

*Preuve du théorème (2.1).*

• **Troncation et coût**

Pour  $R > 0$ , nous définissons  $\Omega_R := [-\pi R, \pi R]^d$  et  $\mu^R := \frac{\mu}{\mu(\Omega_R)} \mathbf{1}_{\Omega_R}$ .

Nous construisons  $X_i^R, i = 1, \dots, n$ , variables aléatoires indépendantes de loi  $\mu^R$  et nous définissons  $\mu_n^R := \sum_{i=1}^n \delta_{X_i^R}$ . Nous avons alors :  $W_1(\mu_n^R, \mu_n) \leq \frac{1}{n} \sum_{i=1}^n Z_i$ , où  $Z_i = 2|X_i| \mathbf{1}_{X_i \notin \Omega_R}$

Et donc

$$\mathbb{E} \left[ W_1(\mu_n^R, \mu_n) \right] \leq 2\mathbb{E}[X^2]^{1/2} \mathbb{P}(X \notin \Omega_R)^{1/2} \quad (2.4)$$

• **Troncation et transformée de Fourier**

Pour tout  $R > 0$  et  $m \in \mathbb{Z}^d$ , nous définissons  $f_{\mu^R}^R(m) := \int_{\Omega_R} e^{i\frac{m}{R}x} d\mu^R(x)$  la transformée de Fourier de  $\mu^R$  sur  $\Omega_R$ .

Si  $u$  est une fonction  $2\pi R$ -périodique, lisse, et Lipschitz de norme 1, avec  $u(x) = \sum_{m \in \mathbb{Z}^d} a_m e^{i\frac{m}{R}x}$ , alors pour tout  $l = 1, \dots, d$  nous avons  $\partial_l u(x) = \frac{i}{R} \sum_{m \in \mathbb{Z}^d} m_l a_m e^{i\frac{m}{R}x}$ .

Puis par l'égalité de Parseval, et en sommant sur  $l = 1, \dots, d$ , nous obtenons

$$\frac{1}{(2\pi R)^d} \int_{\Omega_R} |\nabla u(x)|^2 dx = \sum_{m \in \mathbb{Z}^d} \frac{|m|^2}{R^2} |a_m|^2. \quad (2.5)$$

Ainsi, si  $u$  est Lipschitz de norme de Lipschitz égale à 1, le terme de gauche est borné par 1, et donc

$$R^2 \geq \sum_{m \in \mathbb{Z}^d} |a_m| |m|^2. \quad (2.6)$$

Ainsi, en reprenant exactement le même raisonnement que dans la preuve du théorème (1.3), nous obtenons que pour toutes  $P$  et  $Q$  probabilités à support dans  $[-\pi R, \pi R]^d$ ,

$$W_1(P, Q)^2 \leq R^2 \sum_{m \neq 0} \frac{1}{|m|^2} (f_P^R(m) - f_Q^R(m))^2. \quad (2.7)$$

**. Régularisation et coût**

Pour tout  $R > 0$  et  $x \in \Omega_R$ , nous définissons  $p_t^R(x) := \frac{1}{(2\pi R)^d} \sum_{m \in \mathbb{Z}^d} e^{i \frac{m}{R} x - |m|^2 t}$ .  $p_t^R$  est la densité de  $\gamma_t^R$ , qui est telle que  $f_{\gamma_t^R}^R(m) = e^{-|m|^2 t}$ .

Si  $P$  est une proba sur  $\Omega_R$ , nous définissons  $P_t^R := P * \gamma_t^R$  via l'égalité  $\int_{\Omega_R} g dP_t^R = \int_{\Omega_R} \int_{\Omega_R} g(x+y) p_t^R(y) dy dP(x)$ .

Si  $P \in P(\Omega_R)$  et  $u$  une fonction lisse,  $2\pi R$ -périodique, et 1-Lipschitz, nous avons en utilisant le caractère Lipschitz  $|\int_{\Omega_R} u dP_t - \int_{\Omega_R} u dP| \leq \int_{\Omega_R} |y| p_t^R(y) dy$ . En passant au sup sur les fonctions Lipschitz de norme de Lipschitz égale à 1, nous obtenons :

$$W_1(P_t^R, P) \leq \left( \int_{\Omega_R} |y|^2 p_t^R(y) dy \right)^{1/2}. \quad (2.8)$$

Notons  $M^R(y) := y - 2k\pi R$ , pour  $(2k-1)\pi R \leq y \leq (2k+1)\pi R$ . Cette application pousse en avant une probabilité  $\eta$  sur  $\mathbb{R}$  en une probabilité  $\eta'$  sur  $(-\pi R, \pi R]$  et nous avons  $f_{\eta'}(m) = f_{\eta}(m)$  (car  $e^{i \frac{m}{R} M^R(y)} = e^{i \frac{m}{R} y}$ ). Si nous choisissons pour  $\eta$  une loi gaussienne centrée sur  $\mathbb{R}$  et de variance  $2t$ , nous obtenons  $\eta'$  la marginale une-dimensionnelle de densité  $p_t^R$  sur  $(-\pi R, \pi R]$ . Ainsi  $\gamma_t^R$  est la mesure produit dont les marginales sont l'image d'une gaussienne centrée de variance  $2t$  par l'application  $M^R$ . Par ailleurs, nous avons  $|M^R(y)| \leq |y|$ . Et donc, en intégrant le long de chaque coordonnée, nous obtenons

$$\int_{\Omega_R} |y|^2 p_t^R(y) dy \leq 2dt. \quad (2.9)$$

Puis en combinant (2.8) et (2.9), et en l'appliquant à  $P = \mu_n^R$ , nous obtenons que, uniformément en  $R > 0$  et pour tout  $t > 0$ ,

$$\mathbb{E}[W_1(\mu_{n,t}^R, \mu_n^R)] \leq \sqrt{2dt} \quad (2.10)$$

**. Distance entre les mesures régularisées**

Commençons par remarquer que  $f_{P_t^R}^R(m) = e^{-|m|^2 t} f_P^R(m)$ .

En utilisant l'inégalité (2.7), et en prenant ensuite l'espérance, en se servant de l'inégalité de Jensen et du fait que  $\mathbb{E}[(f_{\mu_n^R}^R(m) - f_{\nu_n^R}^R(m))^2] \leq \frac{4}{n}$ , nous obtenons alors que

$$\mathbb{E}[W_1(\mu_{n,t}^R, \nu_{n,t}^R)] \leq \frac{2R}{\sqrt{n}} \left( \sum_{m \neq 0} \frac{1}{|m|^2} e^{-2|m|^2 t} \right)^{1/2}. \quad (2.11)$$

Nous allons nous servir alors de la même comparaison-série intégrale que nous avons utilisée pour estimer l'inégalité (1.21).

**. Conclusion**

Par inégalité triangulaire, et en utilisant le fait que les  $X_i$  et les  $Y_i$  sont de même loi, nous avons que pour tout  $t > 0$  et  $R > 0$ ,

$$\mathbb{E}[W_1(\mu_n, \nu_n)] \leq 2\mathbb{E}[W_1(\mu_n, \mu_n^R)] + 2\mathbb{E}[W_1(\mu_n^R, \mu_{n,t}^R)] + \mathbb{E}[W_1(\mu_{n,t}^R, \nu_{n,t}^R)] \quad (2.12)$$

en utilisant les équations (2.12), (2.4), (2.10) et (2.11) nous obtenons :

$$\mathbb{E}[W_1(\mu_n, \nu_n)] \leq 4\mathbb{E}[X^2]^{1/2} \mathbb{P}(X \notin \Omega_R)^{1/2} + 2 \left( \frac{2R}{\sqrt{n}} \left( \sum_{m \neq 0} \frac{e^{-2|m|^2 t}}{|m|^2} \right)^{1/2} + 2\sqrt{2dt} \right). \quad (2.13)$$

Il suffit alors d'utiliser une comparaison série-intégrale, et d'optimiser de manière adéquate afin d'aboutir au théorème (2.1). □

## 2.2 Borne inférieure par comparaison avec l'uniforme

Le but de cette partie est de montrer des résultats faciles qui vont nous permettre de comparer les distances considérées pour une loi  $\mu$  avec la distance dans le cas uniforme, et nous servir donc du théorème AKT, démontré pour la première fois dans [AKT84]. Les lois que nous pouvons comparer avec la loi uniforme sont celles dont la loi uniforme est une image Lipschitz. Enonçons tout de suite le théorème que nous démontrons ensuite.

**Théorème 2.2** (Borne inférieure par comparaison avec l'uniforme). *Soit  $d \geq 1$ .*

*Soit  $\lambda$  la loi uniforme sur  $[0, 1]^d$ .*

*Soit  $\mu \in P(\mathbb{R}^d)$  telle que  $\mu = \otimes_{i=1}^d \mu_i$  où  $\forall i = 1, \dots, d, \mu_i \in P(\mathbb{R})$  et  $F_{\mu_i}$  est Lipschitz.*

*Alors,  $\mathbb{E} \left[ W_1(\mu_n, \mu'_n) \right] \geq C \mathbb{E} \left[ W_1(\lambda_n, \lambda'_n) \right]$ , où  $\mu_n, \mu'_n$  sont des mesures empiriques indépendantes associées à la loi  $\mu$ ,  $\lambda_n, \lambda'_n$  des mesures empiriques indépendantes associées à la loi uniforme  $\lambda$ , et où  $C > 0$  ne dépend que des constantes de Lipschitz des  $F_{\mu_i}$ .*

*Preuve du théorème (2.2).* Supposons d'abord que  $d = 1$ .

$\mu$  est donc une probabilité sur  $\mathbb{R}$  dont la fonction de répartition  $F_\mu$  est Lipschitz.

Nous savons que si  $X_1, \dots, X_n$  et  $X'_1, \dots, X'_n$  sont iid de loi  $\mu$ , alors  $F_\mu(X_1), \dots, F_\mu(X_n)$  et  $F_\mu(X'_1), \dots, F_\mu(X'_n)$  sont indépendantes de loi  $\lambda$ .

Notons respectivement  $\nu_n$  et  $\nu'_n$  les mesures empiriques associées aux variables aléatoires  $F_\mu(X_1), \dots, F_\mu(X_n)$  et  $F_\mu(X'_1), \dots, F_\mu(X'_n)$ .

Nous avons alors  $\mathbb{E} \left[ W_1(\lambda_n, \lambda'_n) \right] = \mathbb{E} \left[ W_1(\nu_n, \nu'_n) \right]$ .

Or par dualité de Kantorovich, cette quantité est aussi égale à  $\sup_{f \text{ 1-Lip}} \int f d(\nu_n - \nu'_n)$ .

Or  $\int f d(\nu_n - \nu'_n) = \frac{1}{n} \sum_{i=1}^n (f \circ F_\mu)(X_i) - (f \circ F_\mu)(Y_i)$ .

Puis, comme  $F_\mu$  est supposée Lipschitz,  $f \circ F_\mu$  l'est aussi pour toute fonction  $f$  Lipschitz, et en réutilisant la dualité de Kantorovich nous obtenons bien le résultat annoncé pour  $d = 1$ .

Replaçons nous maintenant dans le cadre du théorème : soit  $d \geq 1$  et soit  $\mu = \otimes_{i=1}^d \mu_i$  où les  $\mu_i$  sont des probabilités sur  $\mathbb{R}$  de fonction de répartition  $F_{\mu_i}$  toutes Lipschitz.

Comme  $\mu$  est la loi produit des  $\mu_i$  et que l'image de  $\mu_i$  par  $F_{\mu_i}$  est la loi uniforme sur  $[0, 1]$ ; si nous notons  $F : (x_1, \dots, x_d) \mapsto \prod_{i=1}^d F_{\mu_i}(x_i)$ , nous obtenons que l'image de  $\mu$  par  $F$  est la loi uniforme sur  $[0, 1]^d$ .

Comme précédemment, en utilisant la dualité de Kantorovich, il nous suffit donc de montrer que  $F$  est Lipschitz.

Notons  $K_i$  la constante de Lipschitz de  $F_{\mu_i}$ . Nous avons alors, pour  $(x_1, \dots, x_d), (y_1, \dots, y_d) \in \mathbb{R}^d$  :

$$\begin{aligned} |F(x_1, \dots, x_d) - F(y_1, \dots, y_d)| &= |F_{\mu_1}(x_1) \dots F_{\mu_d}(x_d) - F_{\mu_1}(y_1) \dots F_{\mu_d}(y_d)| \\ &\leq |F_{\mu_1}(x_1) - F_{\mu_1}(y_1)| |F_{\mu_2}(x_2) \dots F_{\mu_d}(x_d)| \\ &\quad + |F_{\mu_1}(y_1)| |F_{\mu_2}(x_2) \dots F_{\mu_d}(x_d) - F_{\mu_2}(y_2) \dots F_{\mu_d}(y_d)| \\ &\leq K_1 |x_1 - y_1| + |F_{\mu_2}(x_2) \dots F_{\mu_d}(x_d) - F_{\mu_2}(y_2) \dots F_{\mu_d}(y_d)|. \end{aligned}$$

En itérant ce raisonnement on trouve que  $F$  est Lipschitz, de constante de Lipschitz inférieure ou égale à  $\max_{i=1, \dots, d} K_i$ , ce qui conclut. □

### 2.3 Conjecture dans le cas gaussien

Nous nous sommes beaucoup intéressées à l'existence (et à la valeur) de la constante optimale dans le cas de lois gaussiennes, notamment en dimension 2. Pour cela, nous avons essayé d'adapter l'approche "EDP" que l'on retrouve notamment dans les articles [AST19] puis [AG19]. Nous n'avons pas abouti à un résultat significatif mais nous avons tout de même développé quelques pistes que nous présentons ici. La nouveauté par rapport notamment aux articles de M. Ledoux et J.X. Zhu est que l'étape de régularisation se fait avec une sorte "d'Ornstein-Uhlenbeck tronqué" et que nous utilisons une formule de la trace pour les variétés compactes à poids, le théorème (2.3), dont la démonstration est disponible dans [CR19]. Nous donnons ici une conjecture d'une estimée de la limite supérieure dans le cas gaussien en dimension 2 et pour la distance de Wasserstein  $W_2$ . Nous expliquerons pourquoi nous faisons cette conjecture et où sont les endroits où nous avons rencontré des difficultés, et comment nous pourrions éventuellement surmonter ces difficultés.

**Conjecture 2.1.** *Soit  $\mu$  une loi gaussienne standard sur  $\mathbb{R}^2$ . Soient  $(X_i)_{i \geq 0}$  des variables aléatoires iid de loi  $\mu$ . Soit  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ .*

*Alors*

$$\limsup_{n \rightarrow \infty} \frac{n}{(\log n)^2} \mathbb{E} \left[ W_2^2(\mu_n, \mu) \right] \leq 2.$$

Avant de présenter le raisonnement qui nous a conduit à effectuer cette conjecture, présentons les outils que nous avons évoqués juste avant. Par rapport à ce qui a été fait dans la section (2.1) où notre stratégie était d'effectuer des troncatures successives en cubes de plus en plus grand, nous procédons cette fois ci, en nous inspirant du travail de M. Ledoux présenté dans la section (1.5), à des troncations en boules de rayon de plus en plus grand. Cette stratégie semble plus adaptée aux lois gaussiennes qui ont un comportement radial.

Comme annoncé, un outil central que nous souhaitons utiliser afin d'obtenir une borne supérieure explicite, dans le but de pouvoir, à terme, calculer la constante optimale, est d'utiliser une formule de la trace pour les noyaux de la chaleur associés à un laplacien dérivé  $\Delta_f := \Delta - \nabla f \cdot \nabla$ . La formule que nous utilisons provient de [CR19], plus précisément il s'agit du corollaire 3.2 dans l'article, et nous allons nous en servir sur chaque boule de rayon  $R$ , puis nous ferons tendre les rayons des boules vers l'infini, en optimisant en  $R$  grand. Énonçons donc ce théorème.

**Théorème 2.3.** *Soit  $(M, g)$  une variété Riemannienne compacte, soit  $\Delta$  l'opérateur de Laplace-Beltrami sur cette variété. Soit  $f$  une fonction telle que  $\nabla f$  et  $\Delta f \in L^\infty(M, g)$ .*

*Alors, le noyau de la chaleur pour l'opérateur  $L = \Delta + \nabla f \cdot \nabla$  admet l'expansion asymptotique suivante quand  $t \rightarrow 0$  :*

$$\begin{aligned} \text{tr}(e^{-tL}) =_{t \rightarrow 0} & \frac{\text{Vol}(M)}{(4\pi t)^{n/2}} + (4\pi t)^{-n/2} \left[ t \int_M \left( \frac{1}{6} K(z) - \frac{1}{4} |\nabla f(z)|^2 \right) d\mu(z) \right. \\ & \left. + t^2 \left( a_{0,2} - \int_M \frac{1}{6} K(z) V(z) d\mu(z) + \frac{1}{2} \|V\|_{L^2(\mu)}^2 \right) \right] + R(t) \end{aligned}$$

où  $K$  est la courbure scalaire,  $V = \frac{1}{2} \Delta f + \frac{1}{4} |\nabla f|^2$ , et où  $R(t) = o(t^{2-n/2})$ .

Nous ne donnons pas la preuve de ce résultat, qu'il est possible de retrouver dans [CR19]. Cependant faisons quelques remarques sur sa preuve. Celle-ci s'appuie tout d'abord sur une formule de la trace pour le Laplacien usuel sur la variété, que nous avons notamment évoquée dans la preuve du théorème (1.5) (équation (1.35)). La stratégie des auteurs est d'alors de réussir à passer de ce résultat pour le Laplacien à un résultat pour des opérateurs de type Schrödinger  $\Delta_V = \Delta + V$  où  $V$  est un potentiel, en se servant notamment de la formule de Duhamel. Une fois la formule obtenue pour ces opérateurs de Schrödinger, il s'agit de se servir d'une isométrie entre les opérateurs de Schrödinger et les Laplaciens dérivés, isométrie qui conserve la trace de l'opérateur.

Notre stratégie est donc de nous appuyer sur le travail de M. Ledoux en le modifiant quelque peu. Revenons rapidement sur une équation dans la preuve du théorème (1.7) et montrons pourquoi nous voulons utiliser une formule de la trace. Plus précisément regardons le terme de droite dans l'équation (1.56). En développant, en se servant du caractère *iid* des  $X_i^R$ , en utilisant les propriétés de semi-groupe de  $p_t$ , et en optimisant avec  $R \sim \sqrt{\log n}$ , nous obtenons :

$$4 \int_0^\infty \int_M \mathbb{E} [P_s(f_n^{R,t} - 1)] d\mu ds = \frac{1}{2n} \int_{2t}^\infty \int_{\mathbb{R}^d} [p_s(x, x) - 1] d\mu^R(x) ds + O\left(\frac{1}{n^{c'}}\right)$$

pour un  $c' > 1$ .

Nous voyons donc dans le terme de droite apparaître un terme qui ressemble à la trace d'un noyau de la chaleur, puisque nous avons affaire à l'intégrale d'un noyau le long de la diagonale. Cependant, ce n'est pas exactement la trace de l'opérateur, puisque prendre la trace d'un noyau signifie intégrer ce noyau le long de la diagonale contre la mesure invariante associée au semi-groupe; or ici  $\mu^R$  n'est pas la mesure invariante associée au noyau de Mehler. Il faut donc effectuer une étape supplémentaire afin de pouvoir retrouver une formule de la trace et utiliser le théorème (2.3).

De manière schématique, la preuve de M. Ledoux pouvait se décrire de la façon suivante :

$$\mu_n \longrightarrow \mu_n^R \longrightarrow \mu_n^{R,t} \longrightarrow \mu.$$

Nous ajoutons donc une étape pour avoir un schéma de la forme suivante :

$$\mu_n \longrightarrow \mu_n^R \longrightarrow \mu_n^{R,t} \longrightarrow \mu^R \longrightarrow \mu.$$

C'est quand nous allons évaluer la distance entre  $\mu_n^{R,t}$  et  $\mu^R$  que va intervenir la trace, et nous pourrons donc utiliser la théorème (2.3).

Afin d'évaluer l'espérance la distance de Wasserstein  $W_2$  entre  $\mu_n$  et  $\mu_n^R$ , nous réutilisons l'équation (1.50), nous gardons donc en tête pour l'instant que  $R \sim \sqrt{\log n}$ , nous chercherons plus tard à être plus précis dans cette optimisation.

Passons maintenant à l'étape de régularisation. Pour cela, comme annoncé, nous allons régulariser par le noyau associé à la mesure  $\mu^R$  sur chaque boule  $B_R$ . Il faut d'abord clarifier quel semi-groupe de Markov est tel que la mesure gaussienne tronquée est la mesure invariante.

Pour cela, nous définissons  $(P_t^R)_{t \geq 0}$  le semi-groupe de Markov associé au générateur infinitésimal  $L := \Delta - x \cdot \nabla$  sur  $B_R$  avec conditions au bord de Neumann. Montrons que  $\mu^R$  est la mesure invariante associée au semi-groupe  $(P_t^R)_{t \geq 0}$ . Soit  $f$  une fonction suffisamment sympathique, il s'agit de montrer que

$$\int_{B_R} P_t f d\mu^R = \int_{B_R} f d\mu \tag{2.14}$$

Nous avons :

$$\int_{B_R} P_t^R f(x) d\mu^R(x) = \int_{B_R} \int_0^t \frac{d}{ds} P_s^R f(x) ds d\mu^R(x) + \int_{B_R} f d\mu^R \tag{2.15}$$

$$= \int_0^t \int_{B_R} L P_s^R f(x) d\mu^R(x) ds + \int_{B_R} f d\mu^R. \tag{2.16}$$

Nous notons  $g := P_s^R f$  et nous nous intéressons à  $\int_{B_R} Lg d\mu^R(x)$ .

Nous avons  $\mu(B_R) \int_{B_R} Lg d\mu^R(x) = \int_{B_R} (\Delta g - x \cdot \nabla g) e^{-|x|^2/2} dx$ .

Or, si nous notons  $V(x) := e^{-|x|^2/2} \nabla g(x)$ , nous avons

$$\frac{\partial V_i}{\partial x_i}(x) = \left( \frac{\partial^2 g}{\partial x_i^2}(x) - x_i \frac{\partial g}{\partial x_i}(x) \right) e^{-|x|^2/2} \tag{2.17}$$

et donc

$$(\Delta g - x \cdot \nabla g)e^{-|x|^2/2} = \operatorname{div} V(x). \quad (2.18)$$

Ainsi

$$\int_{B_R} (\Delta g - x \cdot \nabla g)e^{-|x|^2/2} dx = \int_{B_R} \operatorname{div} V(x) dx \quad (2.19)$$

$$= \int_{\partial B_R} V(x) \cdot n dx \quad (\text{par la formule de Green}) \quad (2.20)$$

$$= \int_{\partial B_R} e^{-|x|^2/2} \frac{\partial g}{\partial n}(x) dx \quad (2.21)$$

$$= 0 \quad (\text{par conditions au bord de Neumann pour } g = P_s f). \quad (2.22)$$

Finalement nous avons donc bien

$$\int_{B_R} P_t f d\mu^R = \int_{B_R} f d\mu \quad (2.23)$$

ce qui montre que  $\mu^R$  est bien la mesure invariante associée au semi-groupe  $(P_t^R)_{t \geq 0}$ .

Voici donc maintenant la première différence avec les travaux de M. Ledoux, nous définissons  $\mu_n^{t,R}$  par

$$d\mu_n^{t,R} = f_n^{t,R} d\mu \quad (2.24)$$

où

$$f_n^{t,R}(y) := \frac{1}{n} \sum_{i=1}^n p_t^R(X_i^R, y) \quad (2.25)$$

(comme annoncé nous régularisons donc les mesures tronquées par des sortes d'Ornstein-Uhlenbeck tronqués au lieu de régulariser à chaque fois par un Ornstein-Uhlenbeck).

Passons donc maintenant aux étapes qui nous ont fait aboutir à la conjecture (2.1).

*Raisonnement pour la conjecture (2.1).* Par l'inégalité de Young et l'inégalité triangulaire, nous obtenons immédiatement que pour tout  $\alpha > 0$  :

$$\begin{aligned} \mathbb{E}[W_2^2(\mu_n, \mu)] &\leq 4(1 + \alpha^{-1}) \left( \mathbb{E}[W_2^2(\mu_n, \mu_n^R)] + \mathbb{E}[W_2^2(\mu_n^R, \mu_n^{R,t})] + \mathbb{E}[W_2^2(\mu^R, \mu)] \right) \\ &\quad + (1 + \alpha) \mathbb{E}[W_2^2(\mu_n^{R,t}, \mu^R)] \end{aligned} \quad (2.26)$$

Nous isolons  $\mathbb{E}[W_2^2(\mu_n^{R,t}, \mu^R)]$  car c'est de ce terme que va provenir le bon ordre en  $\frac{(\log n)^2}{n}$  dans la preuve du théorème (1.7) donnée par M. Ledoux, et notre intuition est que c'est encore de ce terme que proviendra le bon ordre avec notre nouvelle manière de régulariser, bien qu'il reste encore des difficultés afin de justifier cette intuition.

Il s'agit donc maintenant de majorer un à un les 4 termes dans le membre de droite de l'inégalité.

### . Etape 1

Tout d'abord :

$$\begin{aligned} \mathbb{E}[W_2^2(\mu^R, \mu)] &= W_2^2(\mu^R, \mu) \\ &= \min_{(X,Y), X \sim \mu, Y \sim \mu^R} \mathbb{E}[|X - Y|^2] \\ &\leq \mathbb{E}[|X - X^R|^2] \\ &\leq \mathbb{E}[|X - Z|^2 \mathbf{1}_{|X| > R}] \quad (\text{par définition de } X^R) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}[|X|^2 \mathbb{1}_{|X|>R}] + \mathbb{E}[|Z|^2 \mathbb{1}_{|X|>R}] \text{ (par indépendance de } X \text{ et } Z) \\
&\leq \int_{|x|>R} |x|^2 d\mu(x) + R^2 \mathbb{P}(|X| > R) \\
&\leq C_2 R^2 e^{-R^2/2} + C_2' R e^{-R^2/2}
\end{aligned}$$

Ainsi en prenant  $R = \sqrt{2 \log n}$ , nous obtenons :

$$\mathbb{E}[W_2^2(\mu^R, \mu)] \leq O\left(\frac{\log n}{n}\right). \quad (2.27)$$

. **Etape 2** De manière analogue à l'étape 1,

$$\mathbb{E}[W_2^2(\mu_n, \mu_n^R)] \leq 4 \int_{|x|>R} |x|^2 d\mu(x).$$

Nous obtenons donc, toujours en prenant la même optimisation en  $R > 0$  grand :

$$\mathbb{E}[W_2^2(\mu_n, \mu_n^R)] \leq O\left(\frac{\log n}{n}\right). \quad (2.28)$$

. **Etape 3** Intéressons-nous maintenant à  $\mathbb{E}[W_2^2(\mu_n^R, \mu_n^{R,t})]$ .

Par convexité de  $W_2^2$ , nous avons :

$$\begin{aligned}
W_2^2(\mu_n^R, \mu_n^{R,t}) &\leq \frac{1}{n} \sum_{i=1}^n W_2^2(\delta_{X_i^R}, p_t^R(X_i^R, \cdot)) d\mu^R \\
&= \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}^2} |X_i^R - y|^2 p_t^R(X_i^R, y) d\mu^R(y).
\end{aligned}$$

En prenant l'espérance dans l'inégalité précédente, et en se servant du caractère *iid* des  $X_i^R$ , nous obtenons ainsi :

$$\mathbb{E}[W_2^2(\mu_n^R, \mu_n^{R,t})] \leq \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} |x - y|^2 p_t^R(x, y) d\mu^R(y) d\mu^R(x) \quad (2.29)$$

Voilà le premier endroit où nous rencontrons un problème. Nous aurions en effet besoin d'estimées sur  $p_t^R$  afin d'aboutir à un résultat. Par exemple, dans [Led19b], M. Ledoux utilisait des estimées connues sur le noyau de Mehler pour arriver au résultat suivant (nous gardons ici la même notation mais comme il s'agit d'un résultat de la preuve de M. Ledoux, la régularisation était faite à partir du semi-groupe d'Ornstein-Uhlenbeck sur chaque boule) :

$$\begin{aligned}
\mathbb{E}[W_2^2(\mu_n^R, \mu_n^{R,t})] &\leq \frac{1}{\mu(B_R)^2} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} |x - y|^2 p_t(x, y) d\mu(y) d\mu(x) \\
&\leq \frac{4(1 - e^{-t})}{\mu(B_R)} \\
&\leq 8t \text{ pour } R \text{ suffisamment grand}
\end{aligned}$$

Ainsi en prenant  $t = \frac{1}{2n}$ , M. Ledoux obtenait que

$$\mathbb{E}[W_2^2(\mu_n^R, \mu_n^{R,t})] \leq O\left(\frac{1}{n}\right). \quad (2.30)$$

L'intuition voudrait donc que ce terme ne pose pas problème, même avec notre régularisation, cependant nous n'avons pas réussi à justifier ce point. Nous avons essayé de voir s'il existait des liens entre le noyau de Mehler classique et les noyaux sur la boule  $B_R$  régis par le même opérateur  $L = \Delta - x \cdot \nabla$  avec conditions au bord de Neumann, mais nous n'avons pas trouvé de lien utile.

• **Etape 4**

Reste enfin à évaluer  $\mathbb{E}[W_2^2(\mu_n^{R,t}, \mu^R)]$ .

Nous utilisons tout d'abord le théorème 2 de l'article [Led19b], à savoir l'inégalité (1.53) dans le cas  $p = 2$ , qui donne :

$$W_2^2(\mu_n^{R,t}, \mu^R) \leq 4 \|f_n^{R,t} - 1\|_{H^{-1,2}(\mu^R)} \quad (2.31)$$

$$= 8 \int_0^\infty \int_{B_R} (P_s^R(f_n^{R,t} - 1))^2 d\mu^R ds \quad (2.32)$$

or  $P_s^R(f_n^{R,t} - 1) = \frac{1}{n} \sum_{i=1}^n (P_s^R(p_t^R(X_i^R, \cdot) - 1))$

et par symétrie du semi-groupe, nous avons  $P_s^R(p_t^R(X_i^R, \cdot)(y) = p_{s+t}^R(X_i^R, y)$ .

En combinant cette inégalité avec (2.32), puis en prenant l'espérance, nous obtenons donc :

$$\mathbb{E}[W_2^2(\mu_n^{R,t}, \mu^R)] \leq 8 \int_0^\infty \int_{B_R} \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n (p_{s+t}^R(X_i^R, y) - 1) \right)^2 \right] d\mu^R(y) ds \quad (2.33)$$

Puis par indépendance des  $X_i^R$ , et en utilisant le fait que  $\int_{B_R} p_t^R(x, y) d\mu^R(x) = 1$  pour tout  $t > 0$  et pour tout  $y \in B_R$ , nous remarquons que :

$$\begin{aligned} \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n (p_{s+t}^R(X_i^R, y) - 1) \right)^2 \right] &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} [(p_{s+t}^R(X_i^R, y) - 1)^2] \\ &\quad + \frac{1}{n^2} \sum_{i \neq j} \mathbb{E} [p_{s+t}^R(X_i^R, y) - 1] \mathbb{E} [p_{s+t}^R(X_j^R, y) - 1] \\ &= \frac{1}{n} \mathbb{E} [(p_{s+t}^R(X^R, y) - 1)^2]. \end{aligned}$$

En combinant cette dernière égalité avec (2.33), nous obtenons :

$$\mathbb{E}[W_2^2(\mu_n^{R,t}, \mu^R)] \leq \frac{8}{n} \int_0^\infty \int_{B_R} \mathbb{E} [(p_{s+t}^R(X^R, y) - 1)^2] d\mu^R(y) ds \quad (2.34)$$

$$= \frac{8}{n} \int_0^\infty \int_{B_R} (\mathbb{E} [(p_{s+t}^R(X^R, y))^2] - 1) d\mu^R(y) ds \quad (\text{car } \mathbb{E} [p_{s+t}^R(X^R, y)] = 1) \quad (2.35)$$

$$= \frac{8}{n} \int_0^\infty \int_{B_R} (p_{2(s+t)}^R(y, y) - 1) d\mu^R(y) ds \quad (\text{par symétrie de } p_t) \quad (2.36)$$

$$= \frac{4}{n} \int_{2t}^\infty \int_{B_R} (p_s^R(y, y) - 1) d\mu^R(y) ds \quad (2.37)$$

Notre stratégie est maintenant de décomposer la dernière intégrale selon que  $2t \leq s < T$  ou  $s \geq T$ , avec  $T = T(R) = \log \frac{R^2+1}{R^2-1} \sim \frac{2}{R^2}$ .

Remarquons tout de suite que pour notre choix d'optimisation en  $R \sim \sqrt{\log n}$  et  $t \sim \frac{1}{n}$  nous avons  $T \gg t$ .

Vient alors le second problème dans notre raisonnement. En effet pour  $s$  grand, c'est-à-dire pour  $s \in [T, +\infty)$ , M. Ledoux se servait encore une fois d'estimées connues sur le noyau de Mehler, afin d'arriver à

$$\frac{4}{n} \int_T^\infty \int_{B_R} (p_s(y, y) - 1) d\mu^R(y) ds \leq O\left(\frac{\log n}{n}\right). \quad (2.38)$$

De la même manière que dans la section précédente, nous aurions besoin d'estimées sur  $p_t^R$  afin d'obtenir un résultat similaire. Encore une fois, il semble raisonnable de penser que ce terme ne devrait pas poser trop de problème, mais nous n'avons pas réussi à la justifier.

La dernière étape du raisonnement est d'évaluer l'intégrale (2.37) en  $s > 0$  petit. Pour cela, nous utilisons la formule de la trace pour les variétés à poids (théorème (2.3)). Celle-ci nous donne :

$$\int_{B_R} p_s^R(y, y) d\mu^R(y) = \frac{\text{Vol}(B_R)}{4\pi s} - \frac{1}{16\pi} \int_{B_R} |z|^2 d\mu^R(z) + O(s\text{Vol}(B_R)R^2e^{-R^2/2}). \quad (2.39)$$

En utilisant que  $\text{Vol}(B_R) = \pi R^2$ , nous obtenons :

$$\int_{B_R} p_s^R(y, y) d\mu^R(y) = \frac{R^2}{4s} + \frac{1}{4}(1 - e^{-R^2/2} - \frac{1}{2}R^2e^{-R^2/2}) + O(sR^4e^{-R^2/2}) \quad (2.40)$$

Puis, en intégrant entre  $2t$  et  $T$ , et en utilisant le fait que  $T \sim \frac{2}{R^2}$  nous avons :

$$\begin{aligned} \int_{2t}^{\infty} \int_{B_R} p_s^R(y, y) d\mu^R(y) ds &= \frac{R^2}{4}(\log T - \log 2t) + (T - 2t)\left(\frac{1}{4} - \frac{1}{4}e^{-R^2/2} - \frac{1}{8}R^2e^{-R^2/2}\right) \\ &\quad + O(R^2e^{-R^2/2}) \end{aligned}$$

En prenant  $t = t(n) = \frac{1}{2n}$  et  $R = \sqrt{2 \log n}$ , nous obtenons donc :

$$\int_{2t}^T \int_{B_R} p_s^R(y, y) d\mu^R(y) ds = \frac{(\log n)^2}{2} - \frac{\log n \log \log n}{4} + O\left(\frac{1}{\log n}\right) \quad (2.41)$$

$$= \frac{(\log n)^2}{2} + o((\log n)^2) \quad (2.42)$$

Il s'agirait alors de combiner (2.37), (2.42) et une éventuelle inégalité qui ressemblerait à (2.38) pour obtenir un résultat de la forme :

$$\mathbb{E}[W_2^2(\mu_n^{R,t}, \mu^R)] \leq \frac{2(\log n)^2}{n} + o\left(\frac{(\log n)^2}{n}\right). \quad (2.43)$$

## . Conclusion

Enfin, en combinant (2.26),(2.27),(2.28), et, hypothétiquement des inégalités de la forme de (2.30) et de (2.43), nous obtiendrons l'équation suivante pour tout  $\alpha > 0$  :

$$\limsup_{n \rightarrow \infty} \frac{n}{(\log n)^2} \mathbb{E}[W_2^2(\mu_n, \mu)] \leq 2(1 + \alpha). \quad (2.44)$$

Il suffirait alors de faire tendre  $\alpha$  vers 0 pour obtenir la conjecture (2.1).

Voilà le raisonnement que nous avons suivi pour aboutir à la conjecture (2.1). Comme nous l'avons vu au fil du raisonnement, deux obstacles se dressent face à nous, mais ces deux obstacles semblent en fait être assez similaires : il s'agirait d'obtenir des estimées sur  $p_t^R$ , par exemple une éventuelle comparaison entre  $p_t^R$  et  $p_t$  le noyau de Mehler pour se servir ensuite des estimées utilisées par M. Ledoux dans la preuve de (1.7). □

Par ailleurs, avant de passer à la section suivante sur les questions encore ouvertes, faisons quelques remarques sur la borne inférieure dans le cas gaussien, dans le but d'obtenir une constante optimale. Nous avons aussi essayé d'adapter l'approche de M. Ledoux dans [Led19b], en changeant la méthode de régularisation, et en reprenant notamment le théorème 2 dans [Led19b]. Après quelques calculs, nous arrivons notamment à une expression dans laquelle nous retrouvons, en reprenant les notations utilisées précédemment, un terme faisant apparaître  $p_t(X_i, y) - p_t^R(X_i^R, y)$ . Nous aurions donc encore besoin d'un lien entre  $p_t$  et  $p_t^R$  afin de se débarrasser de ce terme.

### 3 Questions ouvertes et problèmes analogues

Malgré les avancées plus ou moins récentes que nous avons présentées dans la première partie, il reste un grand nombre de questions ouvertes liées plus ou moins directement au théorème AKT.

#### 1. Constante optimale pour $p \neq 2$

Dans la section (1.3), nous avons énoncé le résultat démontré par L. Ambrosio, F. Stra et D. Trevisan concernant la constante optimale dans le cas de la distance de Wasserstein  $W_2$  et de la loi uniforme sur une variété compacte de dimension 1 ou 2. Une question naturelle qui vient alors à l'esprit est de se demander qu'il advient lorsque  $p$  n'est plus égal à 2 mais par exemple quand  $p$  vaut 1 ou même lorsque  $p \in [1, +\infty)$  est quelconque. Nous avons vu dans la section (1.4) que dans le cas où  $p \in [1, +\infty)$  est quelconque en dimension 3 ou plus, la constante optimale existait mais nous ne connaissons pas sa valeur. En dimension 2, l'existence de la constante pour  $p \neq 2$  n'a pas non plus été prouvée, même s'il semble raisonnable au vu des théorèmes (1.5) et (1.6) de penser que celle-ci existe.

#### 2. Constante optimale en dimension $\geq 3$

Nous pouvons aussi nous intéresser à ce qu'il advient en dimension 3 et plus, toujours pour l'instant dans le cas de lois uniformes. Dans leur article [AST19] L. Ambrosio, F. Stra et D. Trevisan affirment que leur démarche ne convient pas aux dimensions supérieures ou égales à 3. Nous avons essayé de reprendre la démarche des auteurs en la modifiant afin d'aboutir à un résultat; cependant cela n'a pas abouti. Encore une fois, il semble raisonnable de penser que cette constante optimale existe, mais rien n'a été prouvé pour le moment. D'ailleurs, l'ansatz des physiciens dans leur article [CLPS14] ne leur permet pas de conjecturer une valeur pour cette constante, et il n'existe pas d'hypothèse à ce sujet.

#### 3. Constante optimale pour des lois générales

Par ailleurs, en faisant le lien avec les sections (1.5) et (2.3), nous pouvons nous demander si la limite existe dans le cas gaussien. Il semble naturel de supposer que oui; cependant le montrer présente des grosses difficultés techniques. En effet, si nous souhaitons par exemple nous inspirer de l'approche de L. Ambrosio, F. Stra et D. Trevisan pour traiter le cas gaussien, l'adaptation est loin d'être aisée. Une des raisons principales est la grande utilisation des propriétés spécifiques des noyaux de la chaleur sur une variété compacte. Dans leur démonstration, les trois auteurs utilisent en particulier grandement l'ultraconductivité du noyau de la chaleur. Or si nous essayons d'adapter la démarche au cas gaussien, nous risquons d'utiliser le noyau de Melher, mais celui-ci ne vérifie pas la propriété d'ultraconductivité! Cela aboutit à des difficultés techniques et une approche différente pourrait être préférable. C'est d'ailleurs ce que nous avons tenté de faire dans la section (2.3) bien que nous n'ayons pas abouti à un résultat significatif. Comme nous le suggérons dans cette section, il semble raisonnable de penser que la limite supérieure pourrait être inférieure ou égale à 2, mais nous aurions besoin pour cela de résultats intermédiaires.

Dans le cas de lois totalement générales, auxquelles nous nous sommes intéressés dans la section (2.1), nous pouvons encore nous demander à quel point le résultat que nous avons obtenu est optimal. Par exemple, comme nous le faisons remarquer dans la section (2.1), notre résultat appliqué aux lois gaussiennes standards ne donne pas le bon ordre de grandeur. Cette perte provient des inégalités de Parseval successives sur des carrés de taille de plus en plus grande, qui fait apparaître un terme proportionnel au côté du carré.

#### 4. Le cas trajectorien

Une question sans doute un peu plus éloignée au premier abord, mais non moins intéressante concerne les espaces de Wiener. Nous nous sommes demandés ce qu'il se passait si au lieu de tirer des points dans l'espace de manière aléatoire, nous jetions des trajectoires, suivant par exemple des trajectoires browniennes.

Formalisons quelque peu cette question : nous notons  $W = C_0([0,1], \mathbb{R})$  l'espace de Wiener des fonctions  $f$  continues de  $[0,1]$  dans  $\mathbb{R}$  telles que  $f(0) = 0$  et notons  $\mu$  la mesure de Wiener sur  $W$ . Enfin nous notons  $H$  l'espace de Cameron-Martin de  $W$ . C'est-à-dire  $H = \{h \in W / h(t) = \int_0^t h'(s) ds, |h|_H^2 := \int_0^1 |h'(s)|^2 ds < +\infty\}$ .  $H$  muni du produit scalaire  $(f, g)_H := \int_0^1 f'(s)g'(s) ds$  est alors un espace de Hilbert. La question du transport optimal, que nous avons essentiellement abordée jusqu'ici pour des mesures sur des domaines de  $\mathbb{R}^d$ , se transpose de manière assez naturelle au contexte de l'espace de Wiener.

Pour le cas des gaussiennes en dimension finie, nous nous sommes grandement servis du semi-groupe d'Ornstein-Uhlenbeck afin de régulariser notre mesure empirique. La mesure de Wiener étant une sorte d'extension des gaussiennes en dimension infinie, il semble légitime de se demander s'il existe un analogue au semi-groupe d'Ornstein-Uhlenbeck dans le cadre de l'espace de Wiener. En fait il existe bien un semi-groupe d'Ornstein-Uhlenbeck en dimension infinie, défini naturellement lorsque nous nous souvenons de sa définition en dimension finie. Cependant celui-ci n'a pas le même effet de régularisation que son analogue en dimension finie et le problème semble bien complexe.

Une autre manière d'aborder le problème dans le cas de trajectoires serait de considérer la norme infinie entre les trajectoires.

Plus précisément nous pouvons nous intéresser à la valeur de

$$\mathbb{E} \left( \inf_{\pi} \frac{1}{n} \sum_{i=0}^n \|X_i - Y_i\|_{\infty} \right) \quad (3.1)$$

où l'infimum est pris sur l'ensemble des permutation  $\pi$  de  $\{1, \dots, n\}$ , et où les  $X_i, Y_i$  sont des variables aléatoires *iid* suivant une loi sur l'espace de Wiener (par exemple des trajectoires browniennes).

Par définition de la norme infinie, la valeur (3.1) est égale à

$$\mathbb{E} \left( \inf_{\pi} \frac{1}{n} \sum_{i=0}^n \sup_{t \in [0,1]} |X_i(t) - Y_i(t)| \right). \quad (3.2)$$

Imaginons que  $t \in [0,1]$  soit fixé. Les  $X_i(t), Y_i(t)$  deviennent ainsi des variables aléatoires *iid* à valeurs dans  $\mathbb{R}$  de loi que nous appelons  $\mu^t$ . Si on suppose de plus que la loi  $\mu$  de départ sur l'espace de Wiener a pour support un ensemble de fonctions uniformément bornées, nous en déduisons que les lois  $\mu^t$  sont à support uniformément borné pour  $t \in [0,1]$ .

Ainsi, toujours à  $t$  fixé, par le théorème "AKT amélioré" de M. Talagrand (1.2), nous obtenons que  $\mathbb{E} \left( W_1(\mu_n^t, \nu_n^t) \right) = O(n^{-\frac{1}{2}})$  où  $\mu_n^t$  et  $\nu_n^t$  sont les mesures empiriques respectivement associées aux  $X_i(t)$  et aux  $Y_i(t)$ .

Ainsi, nous avons

$$\begin{aligned} \inf_{\pi} \frac{1}{n} \sum_{i=0}^n \sup_{t \in [0,1]} |X_i(t) - Y_i(t)| &\geq \inf_{\pi} \sup_{t \in [0,1]} \frac{1}{n} \sum_{i=0}^n |X_i(t) - Y_i(t)| \\ &\geq \sup_{t \in [0,1]} \inf_{\pi} \frac{1}{n} \sum_{i=0}^n |X_i(t) - Y_i(t)| \end{aligned}$$

$$\geq O(n^{-\frac{1}{2}})$$

où la dernière inégalité ne dépend pas de  $t$  car nous avons supposé  $\mu$  à support uniformément borné.

Ainsi obtenons-nous donc que la valeur considérée dans le cadre de trajectoires, avec la norme infinie comme distance, est plus grande que celle de l'espace d'arrivée (en effet nous avons ici fait le raisonnement pour des fonctions à valeurs dans  $\mathbb{R}$  mais tout fonctionne de la même manière pour des fonctions à valeurs dans  $\mathbb{R}^d$ ). Cependant, il est loin d'être évident de prétendre que les ordres de grandeur que nous devrions obtenir sont les mêmes que ceux de l'espace d'arrivée.

## 5. Le théorème de Leighton-Shor

Une dernière question en lien avec ce mémoire est le cas où  $p = \infty$ . Nous nous intéressons donc à la distance de Wasserstein infinie  $W_\infty$  entre une mesure  $\mu$  et sa mesure empirique  $\mu_n$ . En fait, cette question est un théorème qui a été démontré par T. Leighton et P. Shor en 1986, soit seulement deux ans après la première démonstration du théorème AKT, dans [LS89].

La preuve de ce théorème, un peu de la même manière que la preuve originelle du théorème AKT, repose sur des techniques probabilistes et combinatoire. Au vu des démonstrations récentes du théorème AKT, qui se fondent sur des techniques plus analytiques, avec des techniques de régularisation des mesures empiriques, il paraît légitime de se demander si les méthodes appliquées pour résoudre le théorème AKT de manière "contemporaine" ne pourrait pas s'appliquer au théorème de Leighton-Shor. Il s'agirait alors de régulariser la mesure empirique grâce à des noyaux de la chaleur, et de s'intéresser à la distance de Wasserstein  $W_\infty$  entre la mesure régularisée et la mesure de référence.

A priori, au vu des différents résultats, il semble déjà peu probable que le théorème de Leighton-Shor puisse se déduire directement des différentes estimées du théorème AKT, même en utilisant  $W_\infty = \lim_{p \rightarrow \infty} W_p$  ou encore en utilisant une inégalité du type

$W_\infty \leq CW_p^{\frac{p}{p+2}}$  valide pour certaines mesures en dimension 2. En effet dans le cas du théorème AKT en dimension 2, nous observons que le facteur logarithmique et le facteur en  $\frac{1}{n}$  sont élevés à la même puissance, alors que ce n'est pas le cas dans le théorème de Leighton-Shor.

## 4 Appendice

### 4.1 Théorème de Birkhoff-Von Neumann

Comme annoncé en introduction, nous rappelons ici un théorème classique de Birkhoff-Von Neumann qui affirme que les points extrémaux de l'ensemble des matrices bistochastiques sont les matrices de permutation. Commençons tout d'abord par noter que l'ensemble des matrices bistochastiques est un ensemble convexe, et que de manière évidente les matrices de permutation sont des points extrémaux de cet ensemble, le point étant de montrer que ce sont les seules.

**Théorème 4.1** (Théorème de Birkhoff-Von Neumann). *Les matrices de permutation sont les points extrémaux de l'ensemble des matrices bistochastiques.*

De ce théorème, en utilisant le théorème de Krein-Millman, nous en déduisons que l'ensemble des matrices bistochastiques est l'enveloppe convexe de l'ensemble des matrices de permutation. Nous donnons ici une preuve totalement élémentaire du théorème (4.1).

*Démonstration.* Il est facile de voir que les matrices de permutation sont des points extrémaux des matrices bistochastiques; montrons que ce sont les seuls.

Montrons cela par récurrence sur  $n \geq 2$ . Si  $n = 2$ , soit  $A$  une matrice bistochastique.

Nécessairement  $A$  est de la forme :

$$A = \begin{pmatrix} a & 1-a \\ 1-a & a \end{pmatrix},$$

pour un  $a \in [0, 1]$ .

Si  $A$  est une matrice de permutation, alors c'est un point extrémal. Sinon, nous avons  $0 < a < 1$ . Nous pouvons supposer sans perte de généralité que  $0 < a \leq \frac{1}{2}$ .

Dans ce cas, en posant

$$X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

et

$$Y := \begin{pmatrix} 2a & 1-2a \\ 1-2a & 2a \end{pmatrix}$$

, nous avons  $A = \frac{X+Y}{2}$ , avec  $X$  et  $Y$  bistochastiques, et  $A \neq X$ ,  $A \neq Y$ .  $A$  n'est donc pas un point extrémal, et le résultat est donc vrai pour  $n = 2$ .

Si le résultat est vrai pour  $n - 1 \geq 2$ . Montrons qu'il est vrai pour  $n$ .

Soit  $A$  une matrice extrémale de l'ensemble des matrices bistochastiques.

Nous montrons d'abord que  $A$  a au plus  $2n - 1$  coefficients non nuls. Supposons que  $A$  ait au moins  $2n$  coefficients non nuls, que nous notons  $(a_{i_k, j_k})$  avec les couples  $(i_k, j_k)$  tous distincts. Soit  $H$  le sous-espace vectoriel de dimension  $2n$  de  $\mathcal{M}_n(\mathbb{R})$  engendré par les matrices  $E_{i_k, j_k}$ . Par ailleurs, nous notons  $\mathcal{E}_n(\mathbb{R})$  le sous-espace vectoriel de  $\mathcal{M}_n(\mathbb{R})$  engendré par les matrices  $B$  telles que pour tous  $i, j$ ,  $\sum_{k=1}^n b_{i,k} = \sum_{k=1}^n b_{k,j} = 0$ . Il n'est pas difficile de voir que  $\mathcal{E}_n(\mathbb{R})$  est isomorphe à  $\mathcal{M}_{n-1}(\mathbb{R})$  et est donc de dimension  $(n - 1)^2$ . Ainsi, par un argument de dimension, nous avons  $H \cap \mathcal{E}_n(\mathbb{R}) \neq \{0\}$ . Soit donc  $B \neq 0$ ,  $B \in H \cap \mathcal{E}_n(\mathbb{R})$ . Pour  $t \in \mathbb{R}$ , nous notons  $C_t := A + tB$ . Alors pour tout  $(i, j) \neq (i_k, j_k)$ , nous avons  $c_{i,j} = a_{i,j}$ . Par ailleurs, pour  $t$  suffisamment petit, nous avons  $c_{i_k, j_k} = a_{i_k, j_k} + t b_{i_k, j_k} > 0$ . Enfin, comme  $B \in \mathcal{E}_n(\mathbb{R})$ , nous en déduisons que  $\sum_{k=1}^n c_{i,k} = \sum_{k=1}^n a_{i,k} = 1$  et  $\sum_{k=1}^n c_{k,j} = \sum_{k=1}^n a_{k,j} = 1$ . Donc pour  $t$  suffisamment petit et différent de 0,  $C_t$  et  $C_{-t}$  sont bistochastiques, et  $A = \frac{C_t + C_{-t}}{2}$ ,  $A \neq C_t$ ,  $A \neq C_{-t}$ . Donc  $A$  n'est pas extrémale, ce qui contredit l'hypothèse de départ. Donc  $A$  a au plus  $2n - 1$  coefficients non nuls.

Il existe donc nécessairement une ligne d'indice  $i$  avec un seul coefficient  $a_{i,j}$  non nul, et alors  $a_{i,j} = 1$ .  $A$  étant bistochastique tous les coefficients  $a_{k,j}$  avec  $k \neq i$  sont alors nuls. Ainsi la matrice  $A'$  obtenue à partir de  $A$  en supprimant la ligne  $i$  et la colonne  $j$  est une matrice bistochastique de taille  $n - 1$ , et nous vérifions facilement qu'elle est extrémale. Par hypothèse de récurrence  $A'$  est une matrice de permutation, puis  $A$  en est une aussi ce qui conclut.  $\square$

Ce théorème nous permet d'exprimer  $\inf_{\pi} \frac{1}{n} \sum_{i=0}^n \|X_i, Y_{\pi(i)}\|_p$  comme une distance de Wasserstein, que nous allons définir dans la section suivante, entre des mesures empiriques. Les distances de Wasserstein étant des distances provenant du transport optimal, nous avons besoin d'effectuer quelques rappels et de donner des résultats de base de transport optimal, ce que nous faisons tout de suite.

## 4.2 Bases sur le transport optimal

Le principe du transport optimal est le suivant : étant données  $\mu$  une probabilité sur  $X$ ,  $\nu$  une probabilité sur  $Y$  ( $X$  et  $Y$  étant des espaces métriques), et un coût  $c : X \times Y \rightarrow [0, +\infty]$ ;

nous cherchons à connaître la valeur suivante :

$$\inf_{\gamma} \int_{X \times Y} c(x, y) d\gamma(x, y) \quad (4.1)$$

où l'infimum est pris sur les probabilités sur  $X \times Y$   $\gamma$  dont les marginales sont respectivement  $\mu$  et  $\nu$ .

La formulation originale du problème par Gaspard Monge était quelque peu différente. Il s'intéressait à la valeur de

$$\inf_{T \in \mathcal{M}} \int_X c(x, T(x)) d\mu(x) \quad (4.2)$$

où  $\mathcal{M}$  est l'ensemble des fonctions  $T : X \rightarrow Y$  telles que  $T\#\mu = \nu$ . Cependant cette formulation est moins pratique que (4.1) car une telle fonction  $T$  peut très bien ne pas exister, par exemple dans le cas où  $\mu$  est une masse de Dirac et que  $\nu$  est la mesure de Lebesgue. C'est pourquoi nous préférons la formulation (4.1) pour laquelle il existe des résultats d'existence, qui utilisent notamment le théorème d'Arzela-Ascoli, et d'unicité dans certains cas, d'un minimiseur  $\gamma$  de la valeur (4.1), sous certaines conditions sur les espaces  $X$  et  $Y$  (les supposer polonais est par exemple une condition raisonnable), sur les mesures  $\mu$  et  $\nu$  (par exemple en supposer une absolument continue par rapport à la mesure de Lebesgue), et sur le coût  $c$  (par exemple  $c(x, y) = h(x - y)$  avec  $h$  une fonction uniformément convexe); mais nous ne nous attardons pas sur ces résultats et passons plutôt à la définition des distances de Wasserstein  $W_p, 1 \leq p < +\infty$ .

**Définition 4.1** (Distances de Wasserstein). Soit  $1 \leq p < +\infty$ . Soit  $\mu, \nu \in P(\mathbb{R})^d$ .

Nous définissons la distance de Wasserstein  $W_p$  entre  $\mu$  et  $\nu$  par :

$$W_p(\mu, \nu) := \int_{X \times Y} |x - y|^p d\gamma(x, y)$$

où  $\gamma$  réalise l'infimum dans (4.1).

Ainsi, grâce au théorème de Birkhoff-Von Neumann, sommes-nous réduits à considérer la distance de Wasserstein  $W_p$  entre des mesures empiriques. En effet, supposons  $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$  et  $\nu = \frac{1}{N} \sum_{i=1}^N \delta_{y_i}$ . Un couplage de  $\mu$  et  $\nu$  est une matrice bistochastique  $\pi$ .

Le problème de transport optimal entre  $\mu$  et  $\nu$  devient donc

$$\min \left\{ \sum_{i=1}^N c(x_i, y_j) \pi_{i,j} \mid \pi \text{ bistochastique} \right\} \quad (4.3)$$

Le théorème de Birkhoff-Von Neumann implique que ce min est égal à

$$\min \left\{ \sum_{i=1}^N c(x_i, y_{\pi(i)}) \mid \pi \in \mathcal{S}_N \right\} \quad (4.4)$$

Ainsi, en notant  $\mu_n := \frac{1}{n} \sum_{i=0}^n \mu_{X_i}$  et  $\nu_n := \frac{1}{n} \sum_{i=0}^n \mu_{Y_i}$  les mesures empiriques, nous intéressons-nous au comportement asymptotique de :

$$\mathbb{E}(W_p^p(\mu_n, \nu_n)). \quad (4.5)$$

Il existe diverses manières de reformuler le problème (4.1), en essayant d'écrire l'infimum sur  $\gamma$  comme un supremum sur un autre ensemble. Un résultat fondamental de la théorie du transport optimal est justement une reformulation du problème pour la distance  $W_1$ . Il s'agit du théorème de dualité de Kantorovich, que nous utilisons grandement tout au long de ce rapport.

**Théorème 4.2** (Dualité de Kantorovich). Soient  $\mu, \nu \in P(\mathbb{R}^d)$ .

Alors

$$W_1(\mu, \nu) = \sup_{\|f\|_{Lip} \leq 1} \int_{\mathbb{R}^d} f d(\mu - \nu)$$

où  $\|f\|_{Lip} := \sup_{x, y \in \mathbb{R}^d, x \neq y} \frac{|f(x) - f(y)|}{|x - y|}$  est la constante de Lipschitz de  $f$ .

Plus généralement, si  $c : X \times X \rightarrow [0, +\infty]$  est une fonction semi-continue inférieurement, alors (4.1) se reformule de la façon suivante

$$\sup_{(\psi, \phi) \in \Phi_c} \left\{ \int_X \psi d\mu + \int_X \phi d\nu \right\}$$

où  $\Phi_c$  est l'ensemble des fonctions continues bornées telles que pour tous  $x, y \in X$ ,  $\psi(x) + \phi(y) \leq c(x, y)$ .

Nous ne donnons pas de preuve de ces résultats ici, mais nous conseillons d'aller lire [San15] pour trouver les démonstration de ces théorèmes, et pour plus généralement se familiariser à la théorie du transport optimal. Il est bien sûr aussi possible de s'attaquer à [Vil09] mais celui-ci est sans doute moins introductif.

Un autre résultat utilisé assez fréquemment dans les articles que nous avons rencontré notamment dans la preuve du théorème (1.5) est la formule de Benamou-Brenier. Avant d'énoncer ce théorème, nous avons besoin de donner une définition.

**Définition 4.2** (Plan de flot). Etant données deux mesures  $\mu_0$  et  $\mu_1 \in P(M)$ , un plan de flot est la donnée à la fois d'une courbe faiblement continue  $(\mu_t)_{t \in [0,1]} \in C([0,1], P(M))$  et d'un champ de vecteur Borélien dépendant du temps  $(v_t)_{t \in [0,1]}$  tels que, sur  $[0,1] \times M$ , l'équation de continuité suivante soit vraie au sens des distributions :

$$\frac{d}{dt} \mu_t + \operatorname{div}(v_t \mu_t) = 0. \quad (4.6)$$

Nous pouvons alors énoncer la formule de Benamou-Brenier ;

**Théorème 4.3** (Formule de Benamou-Brenier). Etant données deux mesures  $\mu_0$  et  $\mu_1 \in P(M)$ ,

$$W_2^2(\mu_0, \mu_1) = \inf \left\{ \int_0^1 \int_M |v_t|^2 d\mu_t dt : (\mu_t, v_t) \text{ est un plan de flot entre } \mu_0 \text{ et } \mu_1 \right\}$$

Cette fois encore nous ne donnons pas la démonstration de ce résultat, et nous référons à [BB00] où a été donnée la démonstration originale de ce résultat.

### 4.3 Bases sur les semi-groupes de Markov

Une des clés des preuves de [BL19b] et [AST19] est la régularisation des mesures empiriques. Il s'agit de rendre les mesures plus régulières, pour lesquelles il est plus simple d'estimer les distances de Wasserstein. Nous pouvons tout d'abord nous demander pourquoi nous régularisons nos mesures empiriques. En fait, le procédé de régularisation permet d'effacer le mauvais comportement des mesures empiriques à petite échelle. Les mesures empiriques sont en effet des objets dégénérés, et il est plus simple de considérer des objets lisses.

Il existe diverses manières de régulariser une mesure, la plus connue étant sans doute la convolution par une fonction régulière, et c'est celle-ci qu'utilisent S. Bobkov et M. Ledoux dans [BL19b]. En revanche, L. Ambrosio, F. Stra et D. Trevisan utilisent un procédé de régularisation un peu différent à première vue. En effet, leur régularisation est fondée sur les semi-groupes de la chaleur dans les domaines compacts. En fait la notion de régularisation par un semi-groupe de la chaleur est une sorte de généralisation de la convolution : dans  $\mathbb{R}^d$  ces deux notions

coïncident, et la notion de semi-groupe de la chaleur est plus souple car elle se généralise au cadre des variétés Riemanniennes.

Nous avons donc besoin d'introduire quelques objets qui vont dans ce sens, dont les semi-groupes de Markov et les semi-groupes de la chaleur. Avant de donner la définition des semi-groupes de Markov, nous avons besoin de la notion de mesure invariante pour une famille d'opérateurs.

**Définition 4.3** (Mesure invariante). Etant donnée  $\mathbf{P} = (P_t)_{t \geq 0}$  une famille d'opérateurs. Une mesure  $\mu$  est dite invariante pour  $\mathbf{P}$  si pour toute fonction mesurable positive bornée  $f : E \rightarrow \mathbb{R}$ ,

$$\int_E P_t f d\mu = \int_E f d\mu. \quad (4.7)$$

**Définition 4.4** (Semi-groupes de Markov). Soit  $\mathbf{P} = (P_t)_{t \geq 0}$  une famille d'opérateurs définis sur un ensemble de fonctions mesurables à valeurs réelles, et soit  $\mu$  une mesure invariante pour  $\mathbf{P}$ .

$\mathbf{P}$  est un semi-groupe de Markov s'il vérifie les conditions suivantes :

1. Pour tout  $t \geq 0$ ,  $P_t$  est un opérateur linéaire qui envoie des fonctions mesurables bornées sur des fonctions mesurables bornées.
2.  $P_0 = \text{Id}$  (condition initiale).
3.  $P_t(\mathbf{1}) = \mathbf{1}$  où  $\mathbf{1}$  est la fonction constante égale à 1 (conservation de la masse).
4. Si  $f \geq 0$ , alors  $P_t f \geq 0$  (positivité).
5. Pour tous  $t, s \geq 0$ ,  $P_{t+s} = P_t \circ P_s$  (propriété de semi-groupe).
6. Pour toute fonction  $f \in L^2(\mu)$ ,  $P_t f$  converge vers  $f$  dans  $L^2(\mu)$  quand  $t \rightarrow 0$  (propriété de continuité).

Puisque  $(P_t)_{t \geq 0}$  est un semi-groupe, la propriété de continuité affirme que  $t \mapsto P_t f$  est continue.

La notion de semi-groupe de Markov est très prolifique dans le sens où elle engendre et est en lien avec un grand nombre d'autres notions. Elle s'inscrit tout d'abord dans une théorie plus large, celle des opérateurs bornés, avec notamment toute la théorie de Hille-Yoshida. Nous pouvons donc appliquer un grand nombre de résultats qui s'inscrivent dans cette théorie plus générale pour les semi-groupes de Markov. C'est ainsi que nous arrivons en particulier sur la notion de générateur infinitésimal. Si  $(P_t)_{t \geq 0}$  est un semi-groupe de Markov, nous notons  $\mathcal{B} = L^2(\mu)$  où  $\mu$  est la mesure invariante associée au semi-groupe. La théorie de Hille-Yoshida affirme alors qu'il existe un sous-espace vectoriel dense de  $\mathcal{B}$ , appelé le domaine  $\mathcal{D}$  du semi-groupe  $(P_t)_{t \geq 0}$ , sur lequel la dérivée en  $t = 0$  de  $P_t$  existe.

L'opérateur qui envoie une fonction  $f \in \mathcal{D}$  sur cette dérivée  $Lf \in \mathcal{B}$  de  $P_t f$  en  $t = 0$  est appelé le générateur infinitésimal du semi-groupe.

Le domaine  $\mathcal{D} = \mathcal{D}(L)$  est aussi appelé le domaine du générateur, et la connaissance de  $L$  et de  $\mathcal{D}(L)$  caractérise complètement l'action du semi-groupe sur  $\mathcal{B}$ .

La linéarité des opérateurs  $(P_t)_{t \geq 0}$ , allié à la propriété de semi-groupe, montre que  $L$  est la dérivée de  $P_t$  en tout temps. Plus précisément, pour  $t, s > 0$ ,

$$\frac{1}{s}[P_{t+s} - P_t] = P_t \left( \frac{1}{s}[P_s - \text{Id}] \right) = \left( \frac{1}{s}[P_s - \text{Id}] \right) P_t.$$

Puis, en laissant  $s$  tendre vers 0, nous obtenons :

$$\partial_t P_t = L P_t = P_t L. \quad (4.8)$$

Les opérateurs de Markov tels que nous les avons définis, peuvent être représentés par des noyaux, qui correspondent aux probabilités de transition du processus de Markov associé au semi-groupe. Plus précisément, pour toute fonction  $f$  bornée,

$$P_t f(x) = \int_E f(y) p_t(x, dy) \quad (4.9)$$

où, pour tout  $t \geq 0$ ,  $p_t(x, dy)$  est un noyau de probabilité (c'est-à-dire que pour tout  $x \in E$ ,  $p_t(x, \cdot)$  est une mesure de probabilité, et que pour tout ensemble mesurable  $A$  dans la  $\sigma$ -algèbre,  $x \mapsto p_t(x, A)$  est une application mesurable).

Cette représentation par un noyau de probabilité n'est cependant pas automatique, elle nécessite de travailler sur un espace suffisamment sympathique (par exemple un espace polonais), nous faisons référence à [BGL14] (proposition 1.2.3).

Assez souvent, cette famille de noyaux  $p_t(x, dy)$  admet des densités par rapport à une mesure de référence (par exemple, mais pas forcément, la mesure invariante du semi-groupe).

**Définition 4.5** (Noyau de densité). Un semi-groupe de Markov  $(P_t)_{t \geq 0}$  sur  $(E, \mathcal{F})$  est dit admettre des noyaux de densité par rapport à une mesure de référence  $\sigma$ -finie  $m$  sur  $\mathcal{F}$  si, pour toute fonction bornée  $f : E \rightarrow \mathbb{R}$ , et pour  $m$ -presque tout  $x \in E$ ,

$$P_t f(x) = \int_E f(y) p_t(x, y) dm(y).$$

Encore une fois, sous réserve de bonnes propriétés de l'espace ambiant et du semi-groupe, il est possible de prouver l'existence d'un tel noyau de densité (proposition 1.2.5 dans [BGL14]).

Beaucoup de propriétés du semi-groupe  $(P_t)_{t \geq 0}$  peuvent être transférées en des propriétés sur ces noyaux de densité.

Par exemple, si  $L$  est le générateur du semi-groupe, en partant de l'équation (4.8), nous pouvons voir que la fonction  $p_t(x, y)$  pour  $t > 0$  et  $x \in E$  est une solution de l'équation de la chaleur associée à  $L$  :

$$\partial_t p_t(x, y) = L_x p_t(x, y), \quad p_0(x, y) dm(y) = \delta_x. \quad (4.10)$$

Certains semi-groupes de Markov sont plus "remarquables", ou du moins plus utilisés, que d'autres. C'est par exemple le cas du semi-groupe d'Ornstein-Uhlenbeck, dont se servent M. Ledoux et J. X. Zhu dans [Led19a], [Led19b] et [LZ19], dont la mesure invariante est la loi gaussienne standard, que nous définissons maintenant.

**Définition 4.6** (Semi-groupe d'Ornstein-Uhlenbeck). Soit  $\mu$  la loi gaussienne standard sur  $\mathbb{R}^d$ ,  $d \geq 1$ . Si  $f$  est une fonction, nous définissons

$$P_t f(x) := \int_{\mathbb{R}^d} f(e^{-t}x + \sqrt{1 - e^{-2t}}y) d\mu(y) \quad (4.11)$$

La famille  $(P_t)_{t \geq 0}$  ainsi définie est le semi-groupe d'Ornstein-Uhlenbeck de générateur infinitésimal  $L = \Delta - x \cdot \nabla$ .

Il vient quelques propriétés plus ou moins immédiates sur ce semi-groupe. Il est par exemple possible de voir qu'il admet une représentation intégrale par un noyau, dit noyau de Mehler.

Plus précisément  $P_t f(x) = \int_{\mathbb{R}^d} f(y) p_t(x, y) d\mu(y)$  avec

$$p_t(x, y) = p_t(y, x) = \frac{1}{(1 - e^{-2t})^{d/2}} \exp\left(-\frac{e^{-2t}}{2(1 - e^{-2t})} [|x|^2 + |y|^2 - 2e^t x \cdot y]\right) \quad (4.12)$$

Par ailleurs, comme indiquée avant, la famille  $(P_t)_{t \geq 0}$  forme un semi-groupe de Markov, symétrique dans  $L^2(\mu)$ , et pour lequel la formule suivante d'intégration par parties est vraie : pour toutes fonctions  $f, g$  lisses,

$$\int_{\mathbb{R}^d} f(-Lg) d\mu = \int_{\mathbb{R}^d} \nabla f \cdot \nabla g d\mu. \quad (4.13)$$

La littérature sur ce semi-groupe est très complète, il est possible d'aller voir dans [BGL14], ou même dans [LZ19] dans lequel les auteurs font un inventaire bref mais efficace des principales propriétés de ce semi-groupe.

## 4.4 Bases sur les processus de Poisson

Ici aussi, nous ne faisons que donner des définitions et des propriétés élémentaires sur les processus de Poisson, qui sont utilisés dans [GT20].

Il existe des multitudes de livres à ce sujet, nous référons par exemple à [LP18] pour (beaucoup) plus d'informations à ce sujet.

Rappelons tout d'abord la définition d'une loi de Poisson.

**Définition 4.7** (Loi de Poisson). Une variable aléatoire  $X$  suit une loi de Poisson de paramètre  $\gamma > 0$  si pour tout  $k \in \mathbb{N}$ ,  $\mathbb{P}(X = k) = \frac{\gamma^k}{k!} e^{-\gamma}$ .

Avant de passer à la définition d'un processus de Poisson, nous donnons d'abord celle de processus ponctuel, dont les processus de Poisson sont des cas particuliers.

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité et soit  $(\mathbb{X}, \mathcal{X})$  un espace mesurable quelconque.

Nous définissons  $\mathbf{N}_{<\infty}(\mathbb{X})$  comme l'ensemble des mesures  $\mu$  sur  $\mathbb{X}$  telles que pour tout  $B \in \mathcal{X}$ ,  $\mu(B) \in \mathbb{N}$ .

Nous définissons ensuite  $\mathbf{N} = \mathbf{N}(\mathbb{X})$  comme l'espace des mesures qui peuvent s'écrire comme une somme dénombrable de mesures de  $\mathbf{N}_{<\infty}$ .

Par exemple si  $(x_i)_{i=1,\dots,n}$  sont des points de  $\mathbb{X}$ , alors  $\mu := \sum_{i=1}^n \delta_{x_i}$  est un élément de  $\mathbf{N}$ .

Définissons  $\mathcal{N} = \mathcal{N}(\mathbb{X})$  la tribu engendrée par la collection des sous-ensembles de  $\mathbf{N}$  de la forme  $\{\mu \in \mathbf{N} : \mu(B) = k\}$ ,  $B \in \mathcal{X}$ ,  $k \in \mathbb{N}$ .

Passons donc à la définition d'un processus ponctuel.

**Définition 4.8** (Processus Ponctuel). Un processus ponctuel sur  $\mathbb{X}$  est un élément aléatoire  $\eta$  de  $(\mathbf{N}, \mathcal{N})$ , c'est-à-dire une application mesurable  $\eta : \Omega \rightarrow \mathbf{N}$ .

Par exemple, si  $X_1, \dots, X_n, \dots$  sont des variables aléatoires sur  $\mathbb{X}$  et si  $\kappa$  est une variable aléatoire à valeurs dans  $\mathbb{N}$ , alors  $\eta := \sum_{i=1}^{\kappa} \delta_{X_i}$  est un processus ponctuel.

Nous pouvons maintenant donner la définition des processus de Poisson, qui, comme annoncé précédemment, sont des cas particuliers de processus ponctuels.

**Définition 4.9** (Processus de Poisson). Soit  $\lambda$  une mesure sur  $\mathbb{X}$ . Un processus de Poisson d'intensité  $\lambda$  est un processus ponctuel  $\eta$  sur  $\mathbb{X}$  vérifiant les deux propriétés suivantes :

1. Pour tout  $B \in \mathcal{X}$ ,  $\eta(B)$  suit une loi de Poisson de paramètre  $\lambda(B)$ .
2. Pour tout  $m \in \mathbb{N}$ , et pour tous  $B_1, \dots, B_m \in \mathcal{X}$  deux à deux disjoints, les variables aléatoires  $\eta(B_1), \dots, \eta(B_m)$  sont indépendantes.

Sans rentrer dans les détails, il est possible de montrer l'existence de tels processus (théorème 3.6 dans [LP18]), en faisant appel notamment à un principe dit de superposition.

## Références

- [AG19] L. Ambrosio and F. Glaudo. Finer estimates on the 2-dimensional matching problem. *J. Éc. Polytech., Math.*, 6 :737–765, 2019.
- [AKT84] M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4 :259–264, 1984.
- [AST19] L. Ambrosio, F. Stra, and D. Trevisan. A PDE approach to a 2-dimensional matching problem. *Probab. Theory Relat. Fields*, 173(1-2) :433–477, 2019.
- [BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3) :375–393, 2000.
- [BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Cham : Springer, 2014.
- [BGV07] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Relat. Fields*, 137(3-4) :541–593, 2007.
- [BL19a] S. Bobkov and M. Ledoux. *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*, volume 1259. Providence, RI : American Mathematical Society (AMS), 2019.
- [BL19b] S. G. Bobkov and M. Ledoux. A simple Fourier analytic proof of the AKT optimal matching theorem. 2019.
- [BLM16] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities. A nonasymptotic theory of independence. Corrected paperback edition*. Oxford : Oxford University Press, corrected paperback edition edition, 2016.
- [CLPS14] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro. Scaling hypothesis for the euclidean bipartite matching problem. *Physical Review E*, 90(1), Jul 2014.
- [CR19] N. Charalambous and J. Rowlett. The heat trace for the drifting Laplacian and Schrödinger operators on manifolds. *Asian J. Math.*, 23(4) :539–560, 2019.
- [Gla19] F. Glaudo. On the  $c$ -concavity with respect to the quadratic cost on a manifold. *Non-linear Anal., Theory Methods Appl., Ser. A, Theory Methods*, 178 :145–151, 2019.
- [GT20] M. Goldman and D. Trevisan. Convergence of asymptotic costs for random euclidean matching problem. 2020.
- [Led19a] M. Ledoux. On optimal matching of Gaussian samples. *J. Math. Sci., New York*, 238(4) :495–522, 2019.
- [Led19b] M. Ledoux. On optimal matching of Gaussian samples II. 2019.
- [LP18] G. Last and M. Penrose. *Lectures on the Poisson process*, volume 7. Cambridge : Cambridge University Press, 2018.
- [LS89] T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica*, 9(2) :161–187, 1989.
- [LZ19] M. Ledoux and J.X. Zhu. On optimal matching of Gaussian samples III. 2019.
- [San15] F. Santambrogio. *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling*, volume 87. Cham : Birkhäuser/Springer, 2015.
- [Tal92] M. Talagrand. The Ajtai-Komlos-Tusnady matching theorem for general measures. In *Probability in Banach spaces, 8 : Proceedings of the eighth international conference, held at Bowdoin College in summer of 1991*, pages 39–54. Boston, MA : Birkhäuser, 1992.
- [Tal14] M. Talagrand. *Upper and lower bounds for stochastic processes. Modern methods and classical problems*, volume 60. Berlin : Springer, 2014.
- [Tal18] M. Talagrand. Scaling and non-standard matching theorems. *C. R., Math., Acad. Sci. Paris*, 356(6) :692–695, 2018.

- [Vil09] C. Villani. *Optimal transport. Old and new*, volume 338. Berlin : Springer, 2009.
- [Yuk92] J. Yukich. Some generalizations of the Euclidean two-sample matching problem. In *Probability in Banach spaces, 8 : Proceedings of the eighth international conference, held at Bowdoin College in summer of 1991*, pages 55–66. Boston, MA : Birkhäuser, 1992.