

Université de Bordeaux

Master 2 : Modélisation Numérique et Calcul Haute Performance

---

# PROBLÈMES INVERSES APPLICATION AU TRANSPORT OPTIMAL

---

Thomas Philibert - Matthieu Nastorg



FIGURE 1 – WASSERSTEIN Barycenters [2]

# Table des matières

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>                                    | <b>3</b>  |
| <b>2</b> | <b>Notions de Transport Optimal</b>                    | <b>4</b>  |
| 2.1      | Transport entre deux densités . . . . .                | 4         |
| 2.1.1    | Densités de probabilités . . . . .                     | 4         |
| 2.1.2    | Transport entre deux densités . . . . .                | 4         |
| 2.2      | Transport optimal et distance de WASSERSTEIN . . . . . | 5         |
| 2.2.1    | Formulation de Benamou-Brenier . . . . .               | 5         |
| 2.2.2    | Formulation de MONGE . . . . .                         | 6         |
| 2.2.3    | Formulation de KANTOROVICH . . . . .                   | 7         |
| 2.2.4    | Conclusion . . . . .                                   | 8         |
| 2.3      | Barycentres de WASSERSTEIN . . . . .                   | 9         |
| <b>3</b> | <b>Méthode de résolution</b>                           | <b>11</b> |
| 3.1      | Régularisation entropique . . . . .                    | 11        |
| 3.2      | Noyau des distances et nouveau problème . . . . .      | 13        |
| 3.3      | Distance de WASSERSTEIN . . . . .                      | 14        |
| 3.3.1    | Présentation théorique . . . . .                       | 14        |
| 3.3.2    | Résultats . . . . .                                    | 16        |
| 3.4      | Barycentres de WASSERSTEIN . . . . .                   | 17        |
| 3.4.1    | Présentation théorique . . . . .                       | 17        |
| 3.4.2    | Résultats . . . . .                                    | 18        |
| <b>4</b> | <b>Conclusion et perspectives</b>                      | <b>23</b> |

# Chapitre 1

## Introduction

Deux problèmes sont dits inverses l'un de l'autre si la formulation de l'un met l'autre en cause. Ainsi un problème inverse consiste à déterminer des causes connaissant des effets. En pratique, nous sommes plus habitués à résoudre des problèmes directs comme la prédiction de l'état futur d'un système, connaissant son état actuel. Cependant, il est également légitime de se poser le problème "inverse" à savoir : peut-on reconstituer l'état passé d'un système connaissant son état actuel ? Grâce aux nombreuses avancées mathématiques, il est souvent possible de répondre à ce type de problématique à l'aide de diverses techniques. La plus importante étant la reformulation du problème sous la forme de la minimisation d'une fonctionnelle d'erreur entre les mesures réelles et synthétiques.

Au travers de ce rapport, nous allons introduire un type de problème inverse dont les domaines d'application sont pléthoriques : le Transport Optimal. Lorsqu'il évoque cette notion, le célèbre mathématicien Cédric VILLANI parle d'un "coup de neuf pour un très vieux problème". En effet, cette théorie a traversé les siècles par le caractère très concret de ses applications et connaît un véritable regain d'intérêt grâce au développement du Machine Learning et de l'Imagerie. Finance, économie, informatique, géographie, météorologie ... Les domaines d'application semblent infinis pour une théorie qui revient "à la mode".

L'objet de ce rapport est d'implémenter une méthode de calcul de la distance et des barycentres de WASSERSTEIN. Notre étude reprend le document [2]. On propose dans un premier chapitre d'introduire les notions fondamentales liées au transport optimal et de définir les différents problèmes. Le second chapitre sera consacré à la méthode de résolution employée et présentera, par la même occasion, nos résultats. En conclusion, nous parlerons des perspectives envisagées grâce à cette méthode, tout particulièrement en vu de notre stage à venir.

# Chapitre 2

## Notions de Transport Optimal

La problème de transport optimal est introduit pour la première fois en 1781 dans un traité intitulé *Mémoires sur les déblais et les remblais* par Gaspard MONGE. Malgré la complexité Mathématiques du problème, son idée peut être résumée en une simple phrase :

**Comment remplir un trou avec un tas de sable, en minimisant le coût pour le déplacer?**

En d'autres termes, on cherche à déplacer un objet d'un endroit à un autre de la façon la plus économe possible. Ce problème du attendre le milieu du XXeme siècle pour être résolu par Leonid KANTOROVICH.

### 2.1 Transport entre deux densités

Pour commencer nous décrivons la notion de transport entre deux densités de probabilités.

#### 2.1.1 Densités de probabilités

Dans le cadre du transport optimal, les fonctions à transporter sont des densités de probabilités. Soit  $\Omega \subset \mathbb{R}^d$  convexe, compact, de mesure non nulle et de frontière  $\partial\Omega$  Lipschitzienne. On considère donc l'espace suivant :

$$P(\Omega) = \left\{ \rho \in \mathcal{L}^2(\Omega) : \rho \geq 0 \text{ et } \int_{\Omega} \rho(x) dx = 1 \right\}$$

#### 2.1.2 Transport entre deux densités

**Définition 2.1** (Transport entre deux densités).

Transporter une densité  $\rho_0 \in P(\Omega)$  sur une autre densité  $\rho_1 \in P(\Omega)$  correspond à trouver :

- Un chemin  $\rho$  de  $P(\Omega)$ , continu en espace et en temps et tel que  $\rho(0) = \rho_0$  et  $\rho(1) = \rho_1$

- Un champ de vitesse  $v(t, x)$  défini sur  $[0, 1] \times \Omega$  tel que pour tout  $t$ ,  $v(t, x) \in \mathcal{L}^2(\rho(t))$  vérifiant l'équation de continuité, ou l'équation de transport :

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0 \\ \rho(t=0) = \rho_0 \\ \rho(t=1) = \rho_1 \end{cases}$$

avec les flux aux bord non sortants :

$$\rho v \cdot \vec{n} = 0 \text{ presque partout sur } [0, 1] \times \partial\Omega \text{ de plus, on doit}$$

De plus, il faut utiliser dans la formulation faible du problème : pour tout  $\phi \in C^\infty([0, 1] \times \Omega)$

$$\int \int_{[0,1] \times \Omega} \rho(\partial \phi + v \cdot \nabla \phi) dt dx = \int_{\Omega} [\rho_1 \phi(1) - \rho_0 \phi(0)] dx$$

**Définition 2.2** (Transport).

Un transport entre  $\rho_0$  et  $\rho_1$  est un couple  $(\rho(t, x), v(t, x))$  solution de l'équation de continuité. L'ensemble des transports  $(\rho, v)$  sera noté :

$$\Pi(\rho_0, \rho_1)$$

## 2.2 Transport optimal et distance de WASSERSTEIN

Au fils des siècles, plusieurs formulations du problème de transport optimal ont vu le jour. On propose dans cette partie de décrire les problèmes les plus connus. On rappelle que l'on cherche à minimiser l'énergie totale de déplacement entre deux densités.

### 2.2.1 Formulation de Benamou-Brenier

Pour chaque  $(\rho, v)$  de  $\Pi(\rho_0, \rho_1)$ , nous allons associer un coût de déplacement. Il est usuel d'utiliser l'énergie cinétique du transport définie par :

$$E(\rho, v) = \frac{1}{2} \int \int_{[0,1] \times \Omega} \rho |v|^2 dt dx$$

Le problème de transport optimal revient donc à minimiser cette fonctionnelle coût. La solution de ce problème d'optimisation définit la distance de WASSERSTEIN. On a :

$$W_2(\rho_0, \rho_1) = \left( \inf_{(\rho, v) \in \Pi(\rho_0, \rho_1)} \frac{1}{2} \int \int_{[0,1] \times \Omega} \rho |v|^2 dt dx \right)^{\frac{1}{2}}$$

La formulation introduite dans ce paragraphe n'est pas celle que nous utiliserons par la suite. Nous emploierons les formulations de MONGE et particulièrement son amélioration : la formulation de KANTOROVICH.

### 2.2.2 Formulation de MONGE

Comme évoqué dans l'introduction, Gaspard MONGE introduit le problème de transport optimal en 1781 dans son traité *Mémoires sur les déblais et les remblais*.

On représente donc le tas de sable par une densité  $\rho_0(x)$ , le trou une densité  $\rho_1$  et le déplacement par une fonction  $T(x)$  qui déplace la quantité de sable présente en  $x$  vers le point  $y = T(x)$ .

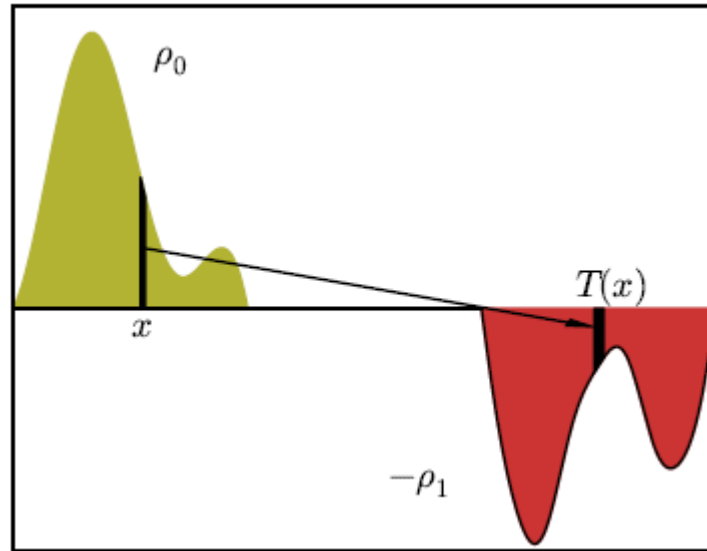


FIGURE 2.1 – Visualisation du problème de MONGE [1]

Le coût du déplacement associé à  $T$  est alors :

$$\int_{\Omega} c(x, T(x)) \rho_0(x) dx$$

Où le terme  $c(x, y)$  est le coût de déplacement unitaire de sable de  $x$  vers  $y$ . Il est usuelle d'utiliser  $c(x, y) = \frac{|x-y|^2}{2}$ .

On cherche donc le déplacement (ou plan de transport)  $T$  optimisant ce coût.

Cependant, nous avons plusieurs contraintes sur les données :

- Le trou et le tas de sable doivent avoir le même volume qu'on fixe à 1. Les fonctions qui les représentent doivent être positives. On a donc que  $\rho_0, \rho_1 \in P(\Omega)$
- On doit de plus déplacer le tas dans le trou ce qui nous donne une contrainte sur le déplacement  $T$ . On écrit cette contrainte par :  $T\#\rho_0 = \rho_1$

En réalité, on ne demande pas à ce que la masse envoyée sur  $T(x)$  soit la même que celle provenant de  $x$ . On écrit plutôt que si  $A \subset \Omega$  est un sous-ensemble quelconque alors tout le sable transporté en  $A$  est exactement le sable qui était en  $T^{-1}(A)$ .

On peut donc l'écrire sous forme mathématique :

$$\forall A \subset \Omega, \quad \int_{T^{-1}(A)} \rho_0(x) dx = \int_A \rho_1(y) dy$$

Si de plus les plans de transport  $T$  sont différentiables, on peut alors effectuer un changement de variable dans la formule précédente et la condition de transport est équivalente à :

$$V = \{T(x), \quad \rho_1(T(x)) | \det(\nabla T)(x) = \rho_0(x)\}$$

Pour résumer, le problème de MONGE est :

$$\begin{aligned} \min_T \quad & \frac{1}{2} \int_{\Omega} |x - T(x)|^2 \rho_0(x) dx \\ \text{s.t} \quad & T \in V \end{aligned}$$

### 2.2.3 Formulation de KANTOROVICH

Au XXème siècle, KANTOROVICH reprend les travaux de MONGE et propose une formulation plus élaborée. C'est celle que nous utiliserons dans le prochain chapitre. Le problème posé est le suivant.

Soient :

- $M$  une variété Riemannienne compacte, connectée et redimensionnée pour avoir un volume unitaire.
- $d : M \times M \rightarrow \mathbb{R}_+$  la distance géodésique sur  $M$  qui est la plus courte distance entre  $x$  et  $y$  sur  $M$ .
- $P(M)$  l'espace de probabilité sur  $M$  et  $P(M \times M)$  la mesure de probabilité sur l'espace produit de  $M$  avec lui-même.

Une densité de probabilité  $\rho_0 \in P(M)$  peut être transporté sur  $\rho_1$  au moyen d'un plan de transport  $\pi \in P(M \times M)$  où  $\pi(x, y)$  décrit le déplacement de la masse de  $\rho_0$  au point  $x$  sur  $\rho_1$  en  $y$ .

La principale différence entre le problème de MONGE et KANTOROVICH est le plan de transport étudié. En effet, dans le problème de MONGE, on ne considère le déplacement qu'entre deux "barrettes" de sable tandis que KANTOROVICH permet de partager la masse d'une "barrette". Dans la figure (2.2), une partie de la masse en  $x$  va en  $y_1$  et l'autre en  $y_2$ . Cela n'était pas possible dans la section précédente.

La loi de conservation des masses impose que ce  $\pi$  soit dans l'ensemble :

$$\Pi(\rho_0, \rho_1) = \{\pi \in P(M \times M) \text{ tel que } \pi(\cdot, M) = \rho_0 \text{ et } \pi(M, \cdot) = \rho_1\}$$

Dans ces conditions, le problème de transport optimal revient à chercher  $\pi \in \Pi(\rho_0, \rho_1)$  tel que le coût soit minimal. Ici il correspond au carré de la distance géodésique .

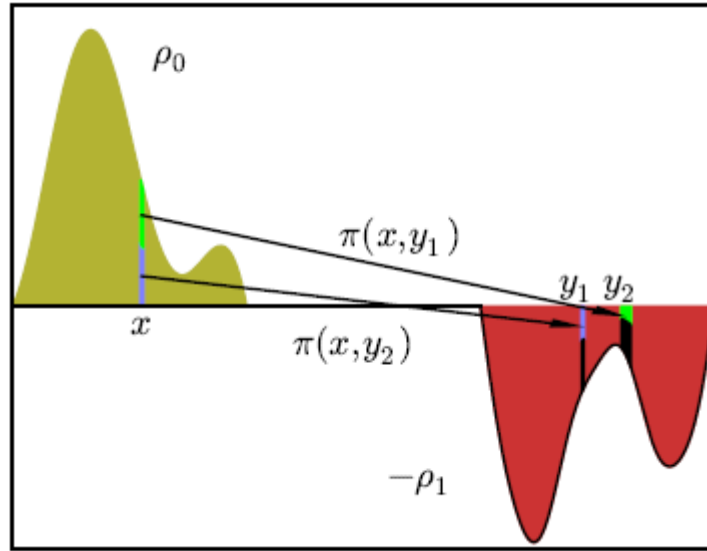


FIGURE 2.2 – Visualisation du problème de KANTOROVICH [1]

Nous avons donc la distance de WASSERSTEIN :

$$W_2(\rho_0, \rho_1) = \left( \inf_{\pi \in \Pi(\rho_0, \rho_1)} \int \int_{M \times M} d(x, y)^2 d\pi(x, y) \right)^{\frac{1}{2}}$$

## 2.2.4 Conclusion

Finalement, nous avons introduit trois formulations du problème du transport optimal. Pour chacun d'eux, le problème se ramène à une étude d'optimisation sous contraintes. Dans le chapitre suivant, nous utiliserons la formulation de KANTOROVICH. Ci-après, nous énonçons un théorème vu dans [1].

### Theorem 2.2.1.

*Les problèmes de MONGE et de BENAMOU-BRENIER sont équivalents, c'est à dire que les valeurs des minimas sont égales (au carré de la distance de WASSERSTEIN).*

La différence entre le problème de MONGE et la formulation de BENAMOU-BRENIER est que l'on cherche un plan de transportation  $T(x)$  à la place d'un chemin  $\rho(t, x)$  et d'un champ de vitesse  $v(t, x)$ . La dimension temporelle a donc disparu. Ce problème, plus simple à écrire, souffre du fait que la contrainte de transport est très difficile à gérer, notamment à cause du déterminant de  $T$ .

La distance de WASSERSTEIN est une métrique sur la position de l'information mais elle n'est pas sensible aux détails de l'information donc à sa forme exacte.



## 2.3 Barycentres de WASSERSTEIN

Dans cette section, on présente la notion de barycentre pour la distance de WASSERSTEIN.

**Définition 2.3** (Barycentre).

Un barycentre de deux ou plusieurs éléments  $x_i$  pour une distance  $d$  est la solution du problème de minimisation :

$$\min_x \sum_i \omega_i d(x, x_i)^2, \quad \omega_i \in \mathbb{R}_+$$

**Définition 2.4** (Barycentre de WASSERSTEIN).

En appliquant la distance de WASSERSTEIN à la définition précédente, on peut définir le barycentre de WASSERSTEIN comme :

$$\min_{\rho} \sum_{i=1}^k \alpha_i W_2^2(\rho, \rho_i)$$

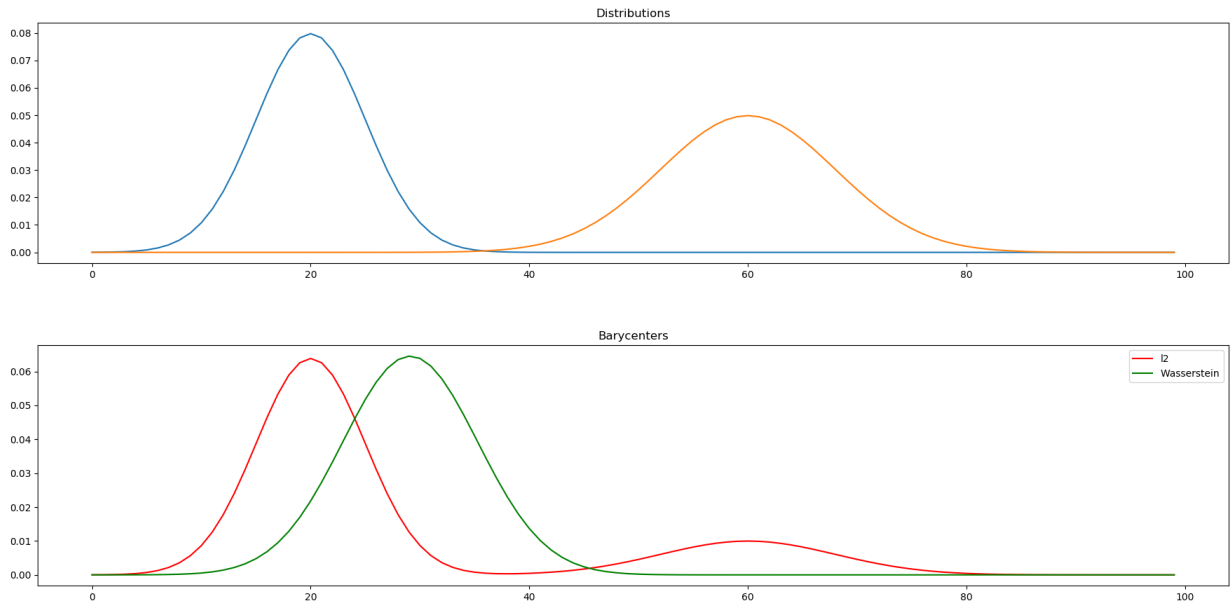


FIGURE 2.3 – Différence entre barycentre de WASSERSTEIN et Euclidien

Le barycentre pour la distance de WASSERSTEIN de deux densités  $\rho_0$  et  $\rho_1$  est une troisième densité dont sa localité dépend des poids  $\omega_0$  et  $\omega_1$ . Ce barycentre a tendance à conserver la forme des densités qu'il considère, ce qui est intéressant dans la mesure où il change la position de ce qu'il interpole.

On peut se demander pourquoi ne pas utiliser le barycentre euclidien plutôt que celui de WASSERSTEIN. Lorsque l'on utilise le barycentre euclidien, on s'aperçoit qu'en réalité, nous faisons une moyenne

pondérée point par point des deux densités qu'on interpole alors que le barycentre de WASSERSTEIN nous permet de transporter une densité sur l'autre.

Voici quelques résultats comparatifs entre les deux barycentres que nous avons créé en utilisant la librairie python que vous pouvez trouver à l'adresse suivante : <https://pot.readthedocs.io/en/stable/>.

Dans la figure 2.3 on trouve en haut nos deux densités  $\rho_0$  et  $\rho_1$ . La partie basse représente le barycentre euclidien en rouge et le barycentre de WASSERSTEIN en vert. On remarque bien dans cet exemple que le barycentre euclidien fait une moyenne pondérée point par point des deux densités alors que celui de WASSERSTEIN va déplacer  $\rho_0$  vers  $\rho_1$

# Chapitre 3

## Méthode de résolution

Nous avons précédemment décrit les notions fondamentales relatives au problème du transport optimal. Cependant, pour réaliser des expériences numériques avec la distance de WASSERSTEIN, encore faut-il savoir la calculer, ce qui n'est pas trivial. La plupart des algorithmes écrits résolvent la formulation continue de BENAMOU - BRENIER :

$$\inf_{(\rho, v) \in C(\rho_0, \rho_1)} \int \int_{[0,1] \times \Omega} \rho \|v\|^2 dt dx$$

puisque, au moyen d'un simple changement de variable, on peut rendre le problème convexe et les contraintes affines (voir [1]). Ainsi, on peut faire appel à une grande catégorie d'algorithmes pour résoudre ce problème d'optimisation sous contraintes. Dans la littérature, on parle, en particulier, d'algorithmes de **descente de gradient** ou bien encore de résolution de problèmes **primal-dual**. Cependant, ces algorithmes peuvent être particulièrement longs à calculer.

Nous nous intéressons, par la suite, à la méthode développée dans [2]. Ce papier, rédigé par nombre de chercheurs reconnus, introduit une nouvelle classe d'algorithmes pour l'approximation du problème de transport optimal sur de **grands domaines géométriques**. Nous proposons de présenter leur méthode pour les cas de dimension 1 et 2. Nous présenterons des résultats pour le calcul de la distance et des barycentres de WASSERSTEIN.

### 3.1 Régularisation entropique

Un facteur limitant pour la résolution du problème de transport optimal est la complexité du problème de minimisation. Pour pallier ce problème, on utilise une méthode de régularisation entropique.

Le principe de cette méthode est le suivant. On considère notre problème initial

$$\min_x f(x)$$

et on le transforme en un nouveau problème auquel on ajoute une petite fonction  $g(x)$ , facile à

minimiser. On parle alors de régularisation du problème de minimisation

$$\min_x (f(x) + \gamma g(x))$$

Pour notre étude, on utilise la formulation de KANTOROVICH vu en (2.2.3).

On rappelle pour commencer le carré de la distance de WASSERSTEIN

$$W_2^2(\rho_0, \rho_1) = \inf_{\pi \in \Pi} \int \int_{M \times M} d(x, y)^2 d\pi(x, y)$$

ainsi que l'espace contraint dans lequel vit  $\pi$

$$\Pi(\rho_0, \rho_1) = \{\pi \in P(M \times M) \text{ tel que } \pi(\cdot, M) = \rho_0 \text{ et } \pi(M, \cdot) = \rho_1\}$$

Puis, on introduit deux outils provenant de la théorie de l'information : la divergence de KULLBACK-LEIBLER et la notion d'entropie d'un plan de transport (cette fonction d'entropie jouera en fait le rôle de la fonction  $g(x)$  dans la régularisation du problème).

**Définition 3.1** (Divergence de KULLBACK-LEIBLER).

Soit  $\mathcal{K}$  une fonction positive sur  $M \times M$ . On définit la divergence de KULLBACK - LEIBLER entre  $\pi$  et  $\mathcal{K}$  comme

$$KL(\pi|\mathcal{K}) = \int \int_{M \times M} \pi(x, y) \left( \ln \frac{\pi(x, y)}{\mathcal{K}} - 1 \right) dx dy$$

La divergence  $KL$  est un outil utilisé en théorie de l'information permettant de comparer deux densités de probabilités. En particulier, elle aide à mesurer la quantité d'information que l'on perd quand on choisit une approximation.

**Définition 3.2** (Entropie d'un plan de transport).

L'entropie d'un plan de transport sur  $M \times M$  est définie comme l'énergie d'entropie concave

$$H(\pi) = - \int \int_{M \times M} \pi(x, y) \ln \pi(x, y) dx dy$$

On modifie le problème de transport optimal en ajoutant le terme d'entropie  $H(\pi)$ . On obtient alors la régularisation entropique de la 2-distance de WASSERSTEIN comme étant le nouveau problème suivant

$$W_{2,\gamma}^2(\rho_0, \rho_1) = \inf_{\pi \in \Pi} \left( \int \int_{M \times M} d(x, y)^2 \pi(x, y) dx dy - \gamma H(\pi) \right)$$

Cette version régularisée du problème de transport optimal est aussi appelé le "problème de SCHRÖDINGER". Quand  $\gamma > 0$  tend vers 0 alors la solution  $\pi$  du problème de minimisation est

absolument continue. Le terme d'entropie  $-H(\pi)$  rend le problème strictement convexe. Par conséquent, selon la théorie développée lors de notre cours de **Problèmes inverses**, il existe un unique minimiseur  $\pi \in \Pi$  du problème. Le problème d'optimisation sous contraintes est bien posé et l'on peut développer une méthode de résolution.

### 3.2 Noyau des distances et nouveau problème

On peut associer la distance  $d(\cdot, \cdot)$  à un noyau  $\mathcal{K}_\gamma$  de la forme

$$\mathcal{K}_\gamma(x, y) = e^{\frac{-d(x, y)^2}{\gamma}}$$

et tel que

$$d(x, y)^2 = -\gamma \ln(\mathcal{K}_\gamma(x, y))$$

On peut alors introduire une nouvelle formulation de la régularisation entropique de la distance de WASSERSTEIN. En effet, cette dernière peut-être comprise comme étant la plus petite  $KL$  divergence d'un plan de transport  $\pi \in \Pi$  par rapport à un noyau  $\mathcal{K}_\gamma$ .

$$W_{2,\gamma}^2(\rho_0, \rho_1) = \gamma \left( 1 + \min_{\pi \in \Pi} KL(\pi | \mathcal{K}_\gamma) \right)$$

Dont la démonstration est

$$\begin{aligned} W_{2,\gamma}^2(\rho_0, \rho_1) &= \inf_{\pi \in \Pi} \left( \int \int_{M \times M} d(x, y)^2 \pi(x, y) dx dy - \gamma H(\pi) \right) \\ &= \inf_{\pi \in \Pi} \left( \int \int_{M \times M} d(x, y)^2 \pi(x, y) dx dy + \gamma \int \int_{M \times M} \pi(x, y) \ln \pi(x, y) dx dy \right) \\ &= \inf_{\pi \in \Pi} \left( \int \int_{M \times M} -\gamma \ln(\mathcal{K}_\gamma(x, y)) \pi(x, y) + \gamma \pi(x, y) \ln(\pi(x, y)) dx dy \right) \\ &= \inf_{\pi \in \Pi} \left( \int \int_{M \times M} \gamma \pi(x, y) (\ln \pi(x, y) - \ln \mathcal{K}_\gamma(x, y)) \right) \\ &= \inf_{\pi \in \Pi} \left( \int \int_{M \times M} \gamma \pi(x, y) \left( \ln \frac{\pi(x, y)}{\mathcal{K}_\gamma(x, y)} \right) \right) \\ &= \inf_{\pi \in \Pi} \left( \int \int_{M \times M} \gamma \pi(x, y) \left( \ln \frac{\pi(x, y)}{\mathcal{K}_\gamma(x, y)} \right) - \gamma + \gamma \right) \\ &= \inf_{\pi \in \Pi} \left( \int \int_{M \times M} \gamma \left( \pi(x, y) \ln \frac{\pi(x, y)}{\mathcal{K}_\gamma(x, y)} + 1 - 1 \right) \right) \\ &= \inf_{\pi \in \Pi} \gamma \left( \int \int_{M \times M} \underbrace{\pi(x, y) \ln \frac{\pi(x, y)}{\mathcal{K}_\gamma(x, y)}}_{KL(\pi | \mathcal{K}_\gamma)} - 1 + 1 \right) \\ &= \gamma \left( \min_{\pi \in \Pi} KL(\pi | \mathcal{K}_\gamma) + 1 \right) \end{aligned}$$

La minimisation est convexe due à la convexité de la divergence de KULLBACK-LEIBLER sur le premier argument  $\pi$  avec des contraintes d'égalité linéaires. Le plan optimal  $\pi$  peut-être vu comme la projection du noyau des distances  $\mathcal{K}_\gamma$  sur  $\Pi$ , tout en imposant les densités  $\rho_0$  et  $\rho_1$  et en minimisant la perte d'information quantifiée par la  $KL$  divergence.

La méthode requiert ici une fonction distance pour construire  $\mathcal{K}_\gamma$ . Pour de grandes variétés arbitraires, calculer la matrice de distance point par point peut s'avérer extrêmement coûteux. La méthode proposée par [2] est de se servir de la résolution de l'équation de la chaleur sur un court intervalle de temps. En effet, le but est de créer une distance  $d(x, y)$  mesurant le transfert de chaleur de  $x$  vers  $y$ . Ainsi le noyau obtenu permet de plus facilement prendre en compte des géométries complexes. Ainsi, plutôt que de sauvegarder la matrice de discrétisation du noyau, il est suffisant de savoir comment l'appliquer à des vecteurs ce qui réduit considérablement le temps de calcul nécessaire. Cependant, comme nous étudions des cas pour des maillages cartésiens avec peu de points, nous utiliserons la matrice de distance euclidienne

$$d(x, y) = |x - y|$$

D'où nous considérerons pour la suite le noyau

$$\mathcal{K}_\gamma(x, y) = e^{-\frac{|x-y|^2}{\gamma}}$$

Il est préférable de sauvegarder la matrice de discrétisation une bonne fois pour toute afin d'éviter de la calculer à chaque fois que l'on lance le code.

### 3.3 Distance de WASSERSTEIN

#### 3.3.1 Présentation théorique

Dans cette section, on s'intéresse au calcul de la distance de WASSERSTEIN. Le but est d'afficher le plan de transport optimal entre deux densités de probabilités  $\rho_0$  et  $\rho_1$ .

Pour faire cela, on considère le travail fait au préalable

$$W_{2,\gamma}^2(\rho_0, \rho_1) = \gamma \left( 1 + \min_{\pi \in \Pi} KL(\pi | \mathcal{K}_\gamma) \right)$$

On implémente nos densités  $\rho_0$  et  $\rho_1$  comme des vecteurs de  $\mathbb{R}^n$  où  $n$  est la taille d'une densité. On considère un vecteur  $a \in \mathbb{R}^n$  tel que  ${}^t a \rho_i = 1$  pour  $i = 0, 1$ . On prendra logiquement  $a = \frac{1}{n}$ . Le plan de transport  $\pi \in P(M \times M)$  est une matrice de taille  $n \times n$  telle que  ${}^t a \pi a = 1$ . La  $KL$  divergence discrète entre un plan de transport discrétisé  $\pi$  et un noyau de distance discrétisé  $H \in \mathcal{M}_{n,n}$  est définie comme

$$KL(\pi | H) = \sum_{i,j} \pi_{i,j} a_i a_j \left[ \ln \frac{\pi_{i,j}}{H_{i,j}} - 1 \right]$$

Finalement on peut modéliser le problème d'optimisation discret, on cherche un plan de transport  $\pi \in \Pi(\rho_0, \rho_1)$  telle que

$$\begin{cases} W_{2, H_t}^2(\rho_0, \rho_1) = \gamma (1 + \min_{\pi \in \Pi} KL(\pi | H_t)) \\ \pi a = \mu_0 \\ {}^t \pi a = \mu_1 \end{cases}$$

En fait, notre matrice de plan de transport  $\pi$  contrainte par son espace  $\Pi(\rho_0, \rho_1)$  possède les propriétés suivantes :

- La somme des lignes de  $\pi$  doit être égale à la première densité
- La somme des colonnes de  $\pi$  doit être égale à la seconde densité.

En fait, on peut comprendre ce problème comme approximer  $\mathcal{K}_\gamma$  avec la matrice de plan de transport  $\pi$  de la façon la plus "proche" que possible, contraint par les propriétés que l'on se doit de respecter entre les lignes et les colonnes de  $\pi$ . On peut voir cela comme un problème de **projection**.

Pour la résolution, les auteurs utilisent l'algorithme de SINKHORN. Ce dernier résout le problème de minimisation en projetant la solution sur chaque contrainte l'une après l'autre. En partant de  $H_t$ , la discrétisation du noyau  $\mathcal{K}_\gamma$ , on projette sur la première contrainte en normalisant les lignes de  $\pi$ . Puis on projette sur la seconde contrainte en normalisant les colonnes de  $\pi$ . En recommençant ce procédé un certain nombre de fois, il a été démontré que cet algorithme converge.

On peut résumer cet algorithme par le croquis ci-dessous

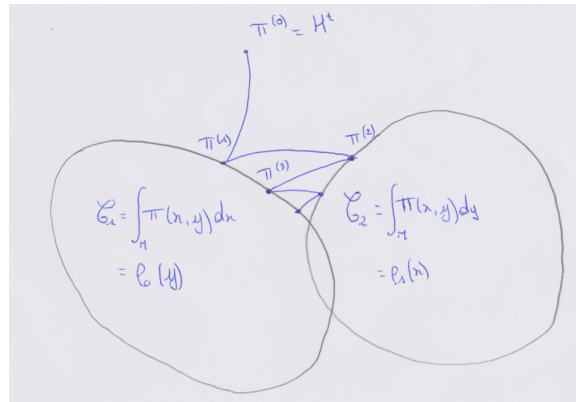


FIGURE 3.1 – Croquis algorithme SINKHORN

Plutôt que de retourner la matrice  $\pi$ , on peut calculer une paire de vecteurs  $(v, w)$  réduisant le nombres d'inconnus à  $2n$ . On détermine  $(v, w)$  en alternant des projections sur les contraintes linéaires. On retrouve  $\pi$  en calculant

$$\pi = D_v H_t D_w$$

où  $D_v$  est la matrice avec  $v$  sur sa diagonale.

Nous avons donc utilisé l'algorithme issu de cette théorie, présenté dans [2]

```

function CONVOLUTIONAL-WASSERSTEIN( $\mu_0, \mu_1; \mathbf{H}_t, \mathbf{a}$ )
  // Sinkhorn iterations
   $\mathbf{v}, \mathbf{w} \leftarrow \mathbf{1}$ 
  for  $i = 1, 2, 3, \dots$ 
     $\mathbf{v} \leftarrow \mu_0 \otimes \mathbf{H}_t(\mathbf{a} \otimes \mathbf{w})$ 
     $\mathbf{w} \leftarrow \mu_1 \otimes \mathbf{H}_t(\mathbf{a} \otimes \mathbf{v})$ 

  // KL divergence
  return  $\gamma \mathbf{a}^\top [(\mu_0 \otimes \ln \mathbf{v}) + (\mu_1 \otimes \ln \mathbf{w})]$ 

```

FIGURE 3.2 – Algorithme pour la distance de WASSERSTEIN [2]

### 3.3.2 Résultats

Nous avons testé l'algorithme pour calculer le plan de transport optimal entre deux densités de probabilités. Nous prenons deux gaussiennes. La première est centrée en 0.2, la seconde en 0.8. On s'intéresse à l'évolution du plan de transport optimal lorsque l'on diminue la valeur de  $\gamma$ .

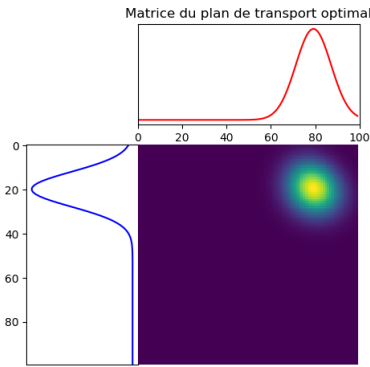


FIGURE 3.3 – PDT  $\gamma = 0.1$

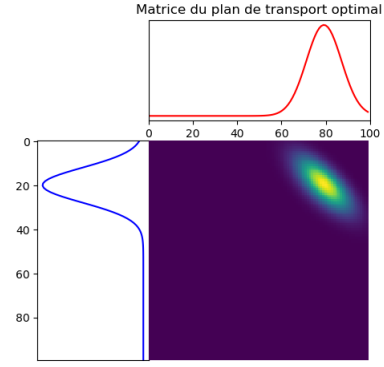


FIGURE 3.4 – PDT  $\gamma = 0.01$

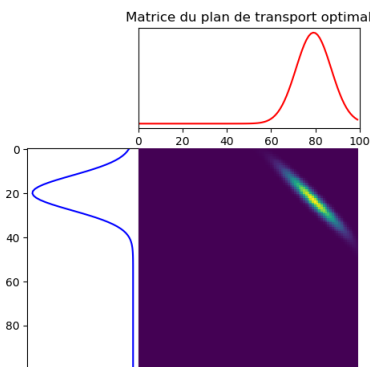


FIGURE 3.5 – PDT  $\gamma = 0.001$

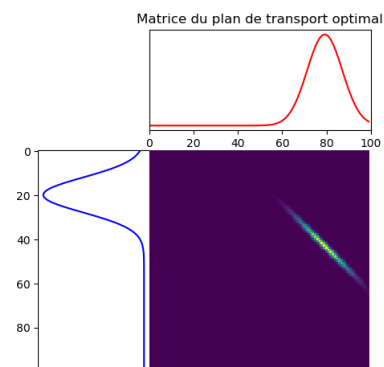


FIGURE 3.6 – PDT  $\gamma = 0.0001$



Comme on peut s'y attendre, lorsque  $\gamma$  est grand, nous avons une perte d'information. Plus  $\gamma$  diminue, plus nous gagnons en précision. Pour un  $\gamma$  de l'ordre de  $10^{-4}$  on perd en justesse d'approximation (on se rend compte que la figure en bas à droite se décale). Cela est dû au noyau de distance que l'on utilise. L'exponentielle rend les valeurs très petites et il est, par conséquent, trop diffusif pour nos applications "simples". C'est un problème récurrent que nous retrouverons par la suite. Une idée serait de trouver un autre noyau (on sait qu'il en existe un mais nous n'avons pas réussi à le trouver dans la littérature) utilisant la fonction logarithmique par exemple.

### 3.4 Barycentres de WASSERSTEIN

Dans cette section, nous détaillons la méthode pour le calcul des barycentres de WASSERSTEIN. Un des intérêts principaux de cette étude est la capacité à calculer ses barycentres pour plusieurs densités de probabilités. Appliqué dans un cadre de la dimension 3, on peut obtenir des résultats d'interpolation comme le montre la figure 1 sur notre page de garde.

#### 3.4.1 Présentation théorique

Le problème du barycentre de WASSERSTEIN vise à moyenner une collection de densités de probabilité  $(\rho_i)_{i=1}^k$  en prenant leur moyenne pondérée par un ensemble de poids en fonction de la distance de WASSERSTEIN.

Soit un ensemble de poids  $\alpha = (\alpha_i)_{i=1}^k \in \mathbb{R}_+^k$ , le problème du barycentre de WASSERSTEIN est défini comme le problème convexe sur l'espace des mesures suivant :

$$\min_{\rho} \sum_{i=1}^k \alpha_i W_2^2(\rho, \rho_i)$$

Après discrétisation, on peut poser ce problème de la façon suivante :

$$\min_{\rho} \sum_{i=1}^k \alpha_i W_{2,H_t}^2(\rho, \rho_i)$$

On obtient le résultat suivant (dont la démonstration nous échappe!) :

$$\begin{aligned} \min_{\pi_i} \quad & \sum_{i=1}^k \alpha_i KL(\pi_i | H_t) \\ \text{s.t.} \quad & \pi_i^T a = \rho_i \quad \forall i \in \{1, \dots, k\} \\ & \pi_i a = \pi_1 a \quad \forall i \in \{1, \dots, k\} \end{aligned}$$

La première contrainte impose que la densité  $\rho_i$  soit la loi marginale de  $\pi_i$  dans une direction et la seconde contrainte impose que un seul  $\rho$  soit la loi marginale de tous les  $\pi_i$  dans la direction

opposée. Suggéré dans [3], le problème peut-être vu comme la projection de  $H_t$  en fonction de la  $KL$  divergence sur l'ensemble des contraintes  $\mathcal{C}^1 \cap \mathcal{C}^2$  où

$$\mathcal{C}^1 = \{(\pi_i)_1^k : \pi_i a = \rho_i \quad \forall i \in \{1, \dots, k\}\} \quad (3.1)$$

$$\mathcal{C}^2 = \{(\pi_i)_1^k : \pi_i a = \pi_j a \quad \forall i, j \in \{1, \dots, k\}\} \quad (3.2)$$

Les auteurs de [2] s'appuie sur la méthode *iterated Bregman projection* pour développer l'algorithme suivant, que nous avons utilisé.

```

function WASSERSTEIN-BARYCENTER( $\{\mu_i\}, \{\alpha_i\}; H_t, \mathbf{a}$ )
  // Initialization
   $\mathbf{v}_1, \dots, \mathbf{v}_k \leftarrow \mathbf{1}$ 
   $\mathbf{w}_1, \dots, \mathbf{w}_k \leftarrow \mathbf{1}$ 

  // Iterate over  $\mathcal{C}_i$ 's
  for  $j = 1, 2, 3, \dots$ 
     $\mu \leftarrow \mathbf{1}$ 
    for  $i = 1, \dots, k$ 
      // Project onto  $\mathcal{C}_1$ 
       $\mathbf{w}_i \leftarrow \mu_i \odot H_t(\mathbf{a} \otimes \mathbf{v}_i)$ 
       $\mathbf{d}_i \leftarrow \mathbf{v}_i \otimes H_t(\mathbf{a} \otimes \mathbf{w}_i)$ 
       $\mu \leftarrow \mu \otimes \mathbf{d}_i^{\alpha_i}$ 

    // Optional
     $\mu \leftarrow \text{ENTROPIC-SHARPENING}(\mu, H_0; \mathbf{a})$ 

    // Project onto  $\mathcal{C}_2$ 
    for  $i = 1, \dots, k$ 
       $\mathbf{v}_i \leftarrow \mathbf{v}_i \otimes \mu \odot \mathbf{d}_i$ 

  return  $\mu$ 

```

FIGURE 3.7 – Algorithme barycentre WASSERSTEIN dans [2]

### 3.4.2 Résultats

Nous présentons dans cette parties nos résultats sur le calcul des barycentres de WASSERSTEIN entre deux densités de probabilité. On s'intéressera aux dimensions 1 et 2 ainsi qu'à l'évolution des solutions en fonction de la valeur de  $\gamma$ .

#### 3.4.2.1 Dimension 1

Dans un premier temps, nous avons appliqué la méthode vue précédemment à deux gaussiennes :  $\rho_0$  centrée en 0.2 et d'écart type 0.08 et  $\rho_1$  centrée en 0.8 et d'écart type 0.08 aussi.

Dans cette exemple, nous allons utiliser  $\omega_0 = \{0.25, 0.5, 0.75\}$  et  $\omega_1 = 1 - \omega_0$ . Nous calculerons les barycentres pour une valeur de  $\gamma = 10^{-4}$ .

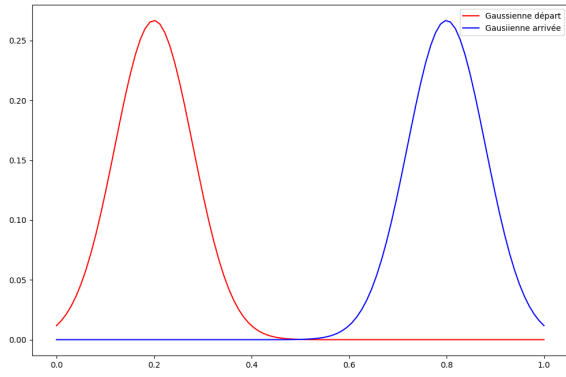


FIGURE 3.8 – Gaussienne  $\rho_0$  en rouge  $\rho_1$  en bleu

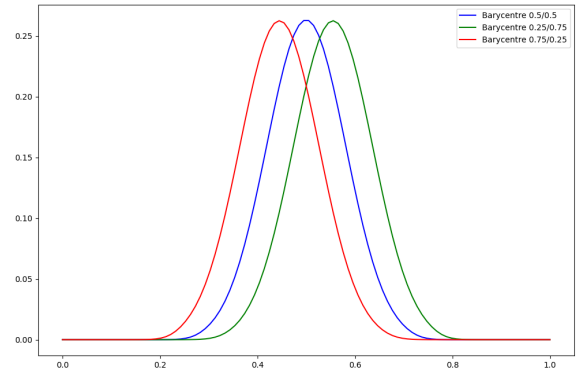


FIGURE 3.9 – Barycentre avec  $\omega_0 = \{0.25, 0.5, 0.75\}$

Si l'on diminue notre  $\gamma$  d'un facteur 10, i.e on applique la méthode pour  $\gamma < 10^{-5}$  on obtient des erreurs. En fait, nous n'obtenons que des valeurs nulles. Ceci s'explique par le fait que le noyau utilisé dans cette méthode est très diffusif et réduit considérablement les valeurs étudiées :

$$\mathcal{K}_\gamma = e^{-d(x,y)^2/\gamma}$$

On remarque que les interpolations sont des gaussiennes qui se trouvent entre  $\rho_0$  et  $\rho_1$ . De plus lorsque  $\omega_0 = 0$  ou 1 et  $\omega_1 = 1 - \omega_0$ , nous retrouvons bien  $\rho_0$  lorsque  $\omega_0 = 0$  et  $\rho_1$  lorsque  $\omega_0 = 1$

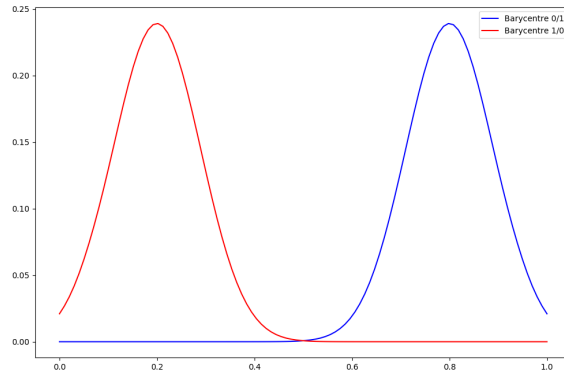


FIGURE 3.10 – Barycentre avec  $\omega_0 = 0$  en rouge et 1 en bleu

Dans ce cas nous avons considéré des gaussiennes "éloignées". Si l'on teste pour plusieurs barycentres, on se rend compte que l'approximation est mauvaise due aux valeurs nulles entre les deux gaussiennes. Cependant, si on les rapproche, on peut afficher avec moins d'erreurs les divers barycentres obtenus (voir figure 3.11)

Finalement, on peut s'intéresser à l'évolution des barycentres en fonction du paramètre  $\gamma$ . On affiche ci-dessous, pour les deux gaussiennes initiales "éloignées", le barycentre à  $\omega_0 = 0.5$  pour diverses valeurs de  $\gamma$ . On portera également un oeil attentif à la norme du barycentre obtenu.

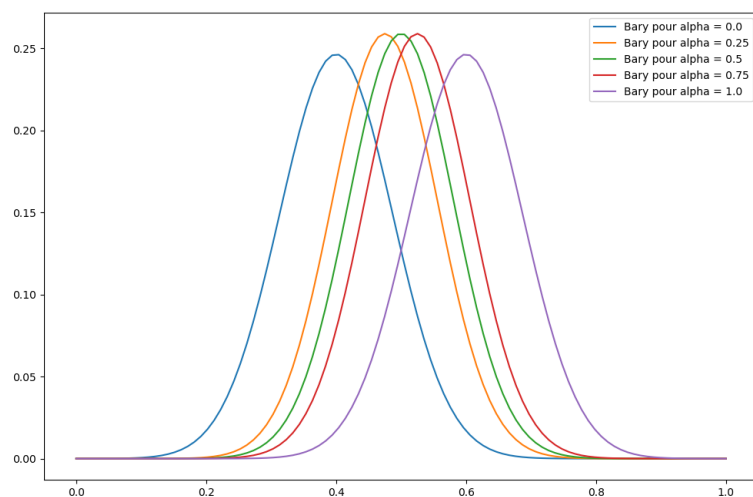


FIGURE 3.11 – Barycentre avec un  $\omega$  évoluant entre 0 et 1

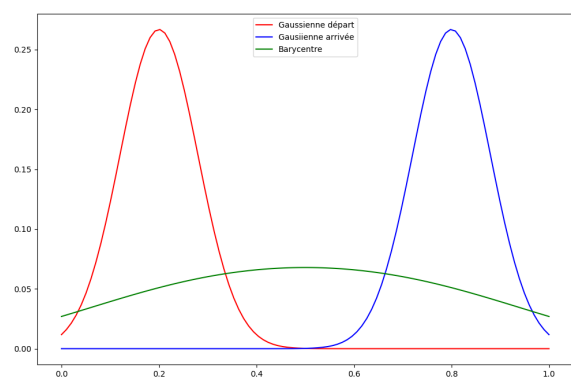


FIGURE 3.12 – Barycentre  $\gamma = 0.1$  et norme = 0.54

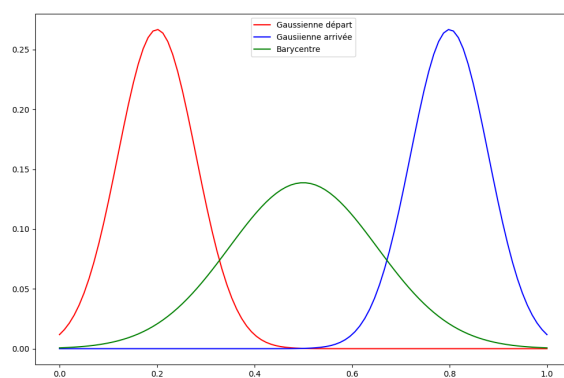


FIGURE 3.13 – Barycentre  $\gamma = 0.01$  et norme = 0.71

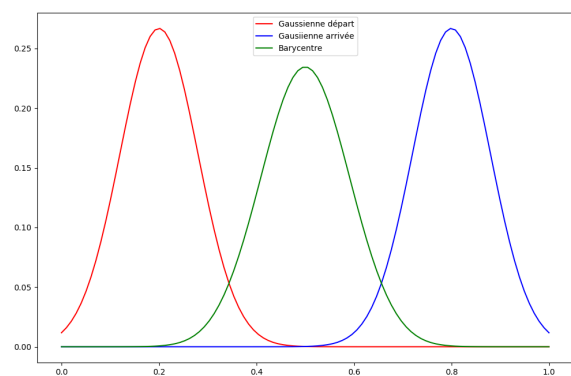


FIGURE 3.14 – Barycentre  $\gamma = 0.001$  et norme = 0.936

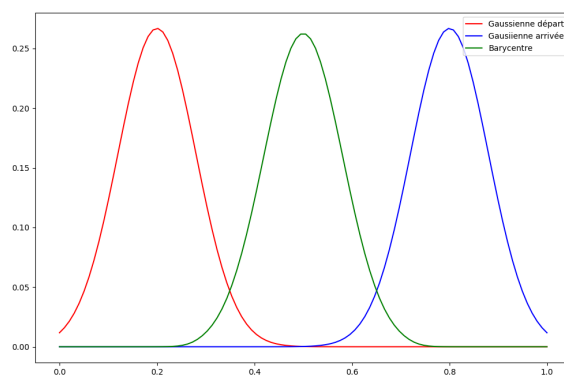


FIGURE 3.15 – Barycentre  $\gamma = 0.0001$  et norme = 0.990

On s'attend à ce que pour un  $\gamma$  raisonnable cette norme soit de 1 (on retrouve une densité de probabilité).

On remarque que, d'après les figures ci dessus, plus  $\gamma$  est petit, plus la norme tend vers 1. Pour les mêmes raisons de diffusivité présentées plus haut, on ne peut descendre en dessous de  $10^{-4}$  pour l'approximation des barycentres. On retrouve donc bien des résultats cohérents ce qui valide la méthode et nous permet de l'étendre à la dimension supérieure.

### 3.4.2.2 Dimension 2

Dans cette partie, nous allons nous intéresser au cas de la dimension 2. Il est important de noter que la matrice de distance à créer devient déjà numériquement très lourde. Par exemple, si nous prenons une discrétisation d'un domaine carré  $[0, 1] \times [0, 1]$  avec un maillage uniforme  $N \times N$  avec  $N = 50$  cela donne  $\mathcal{K}_\gamma \in \mathcal{M}_{2500,2500}$ .

Comme en dimension 1, nous faisons le barycentre de deux gaussiennes. Ici on prendra  $N = 20$ . Nous partons donc avec les gaussiennes  $\rho_0$  et  $\rho_1$  ci dessous :

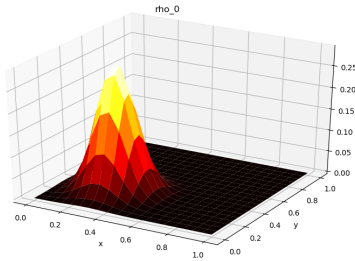


FIGURE 3.16 – Gaussienne  $\rho_0$

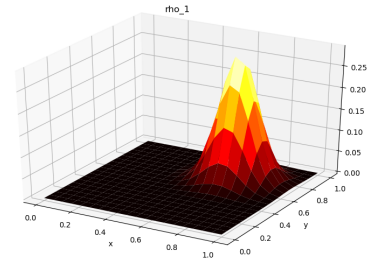


FIGURE 3.17 – Gaussienne  $\rho_1$

En appliquant l'algorithme présenté précédemment avec  $\omega_0 = 0.5$  et  $\gamma = 10^{-3}$ , nous obtenons la courbe 3.18. Si l'on descend sous la barre des  $10^{-3}$  alors on a une grande diffusivité et les résultats refusent de s'afficher.

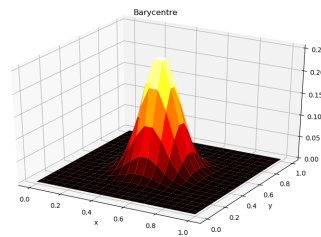


FIGURE 3.18 – Barycentre avec  $\omega_0 = 0.5$ , norme = 0.93

On remarque que la norme de la gaussienne est proche de 1. Finalement, nous présentons un autre style d’affichage. On retrouve sur la figure suivante les deux gaussiennes initiales et leur interpolation pour  $\omega_0 = 0.5$ . On utilise une valeur de  $\gamma = 10^{-3}$ . Nous retrouvons notre gaussienne interpolée entre les deux premières.

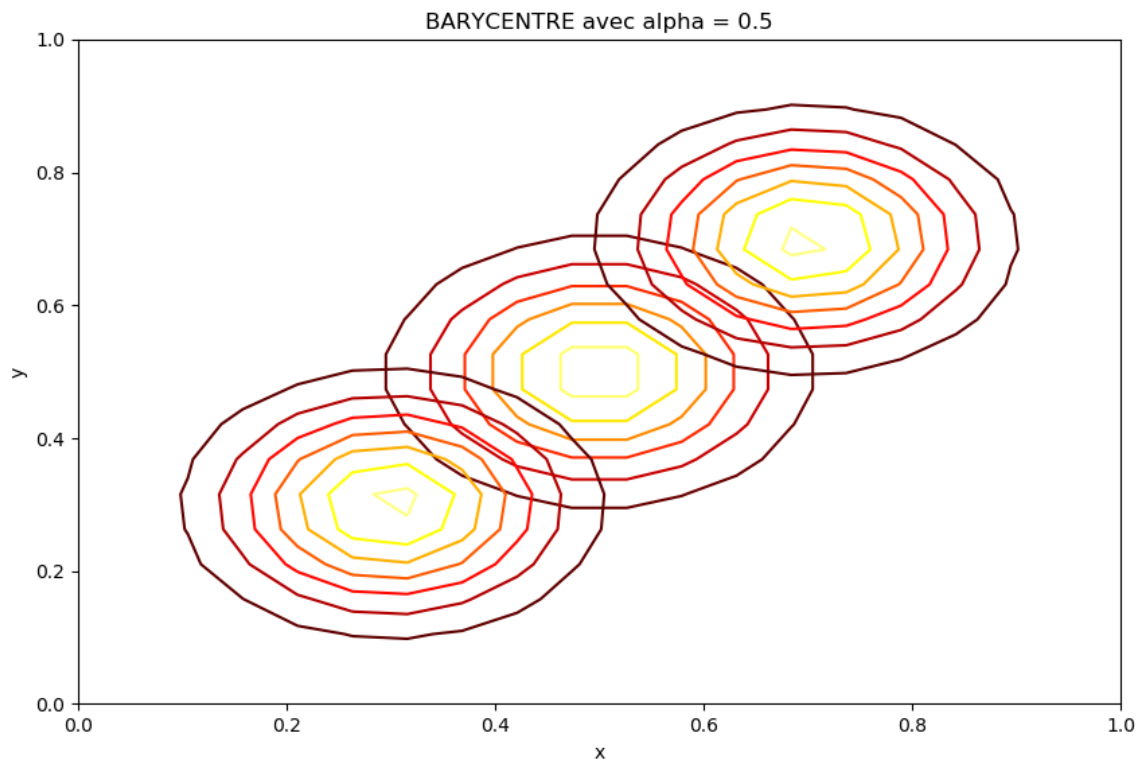


FIGURE 3.19 – Barycentre avec  $\omega_0 = 0.5$

# Chapitre 4

## Conclusion et perspectives

Pour conclure, ce chapitre dénote certains points que nous n'avons pas eu le temps de développer ainsi que des perspectives d'application, en particulier pour notre stage à venir.

Nous avons tout d'abord remarqué la diffusion trop importante de notre noyau des distances. De plus, le temps nécessaire pour calculer cette matrice dans le cas  $2D$  est, pour peu de points, déjà énorme. Cela nous a montré l'importance de la méthode développée dans [2], que nous espérons pouvoir étudier plus en détails par la suite! En effet, nous aurions aimé être en mesure d'appliquer ce résultat sur des squelettes mathématiques d'anévrisme, mais nous avons été bloqué par les deux problèmes précédemment cités (perte d'informations par diffusion numérique et temps de calcul trop élevé ... pour nos ordinateurs).

Certains résultats proposés dans [2], notamment l'utilisation de la *propagation de WASSERSTEIN* serait également intéressant à étudier plus en détails. De plus, nous n'avons pas testé l'utilisation de nos codes pour le barycentre de plus de deux densités.

Finalement, nous aimerions conclure en disant que, parmi les divers sujets que nous avons étudié tout au long de notre Master, le Transport Optimal se révèle être une thématique qui nous a particulièrement plu.

# Bibliographie

- [1] Nelson FEYEUX. “Transport optimal pour l’assimilation de données images”. In : (2016).
- [2] Justin Solomon Fernando de GOES GABRIEL PEYRÉ MARCO CUTURI ADRIAN BUTSCHER ANDY NGUYEN TAO DU LEONIDAS GUIBAS. “Convolutional Wasserstein Distances : Efficient Optimal Transportation on Geometric Domains. (English)”. In : (2015).
- [3] Jean-David Benamou Guillaume Carlier Marco Cuturi Luca Nenna Gabriel PEYRÉ. “Iterative Bregman Projections for Regularized Transportation Problems”. In : (2015).