

Introduction à la théorie du transport optimal

Adil Ahidar*, Aymen Lakhyar*, Badr-Eddine Marani*, Mohamed Amine Hasnabi*, Othmane Mirinioui*, Oumaima Et-Tahri* et Oussama Mouhtal*

*Ecole Centrale Casablanca

Abstract. In this paper, we introduce the theory of optimal transport as it was formulated by Gaspard Monge and Kantorovitch and its formulation in a more general context by measure and probability theories. We discuss some mathematical aspects of this problem in order to introduce the notion of duality and the theorems of existence and characterization of optimal transports.

Therefore, We show how optimal transport theory can be used for a wide range of practical problems, by showing how Optimal Transport problem can allow us to define the Wasserstein distance, which captures the most natural geometry in the space of probability measures. Finally, we present one of these practical problems (MRI (Magnetic Resonance Imagery) paper, also matching and pricing in two-side market.

1. La théorie du transport optimal

1.1. Formulation de Monge

La théorie du transport optimal a été né avec Gaspard Monge en 1781 dans son mémoire "[La théorie des remblais et des déblais](#)" quand il était intéressé par le problème du transport d'un tas de sable vers un trou par des soldats en minimisant l'effort totale du transport.

Formellement, Monge modélise le tas de sable par sa densité de masse $f : \mathbb{R}^d \rightarrow \mathbb{R}$ et le trou cible par la densité de masse $g : \mathbb{R}^d \rightarrow \mathbb{R}$ qu'il peut contenir. Ensuite, Monge considère que le transport de masse du tas vers le trou se fait d'une manière infinitésimale en transportant la masse élémentaire contenu dans chaque point x du trou vers un point y qui pourrait contenir cette masse.

Il s'agit donc de donner une application $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ qui associe à chaque point x le point $y = T(x)$ où la masse contenue dans x sera transporté. Ainsi, la masse transportée dans un endroit \mathcal{B} doit être égale à la masse contenue dans les points $T^{-1}(\mathcal{B})$ d'où cette masse est parvenue, c'est-à-dire :

$$\text{Pour tout borélien } \mathcal{B}; \int_{T^{-1}(\mathcal{B})} f(x) dx = \int_{\mathcal{B}} g(y) dy \quad (1)$$

Cette relation exprime que T transporte la masse répartie selon f à une répartition de masse selon g .

Ensuite, Monge s'intéresse à trouver de tels applications T qui minimisent l'effort totale du transport qui est dans ce cas proportionnel à la distance et à la masse transportée :

$$\inf_{T \text{ vérifie l'équation 1.}} \int_{\mathbb{R}^2} d(x, T(x)) f(x) dx$$

Remarquons qu'un tel transport ne pourra avoir lieu que si la masse totale de la distribution des tas est égale à la masse totale de la distribution cible ($\mathcal{B} = \mathbb{R}^d$ dans 1.) Quitte donc à normaliser par la masse totale, les deux distributions de masses peuvent être vu comme deux mesures de probabilités sur \mathbb{R}^d et le problème de Monge peut être donc se généraliser à n'importe quelles mesures de probabilités μ et ν et l'égalité 1 se traduit par le fait que ν est la mesure image de μ par l'application T . On peut la noter par : $\nu = \mu \# T$.

D'une manière encore plus générale le problème de Monge se formule donc de la manière suivante :

Formulation de Monge (1781). Étant donné deux mesures de probabilités μ et ν sur un espace Ω mesurable et une fonction coût $c : \Omega \times \Omega \rightarrow \mathbb{R}$.

$$(M) : \inf_{\nu = \mu \# T} \int_{\Omega} c(x, T(x)) d\mu$$

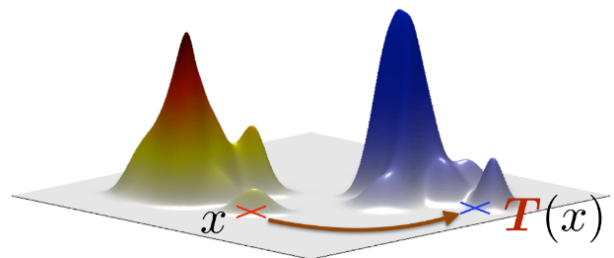


FIGURE 1 | Source : [Cuturi]

Le problème de Monge est donc un problème d'optimisation sous contrainte sur un espace fonctionnel, et comme tout problème d'optimisation des questions naturels se posent :

- Existe toujours des solutions ?
- Les solutions s'elles existent sont-elles uniques ?
- Peut-on caractériser ces solutions ? Et surtout peut on les calculer ?

Malheureusement, dans son mémoire Monge n'a pas pu répondre à ces questions même dans le cas des mesures absolument continues dans l'espace physique et il n'a pu trouver que quelques propriétés géométriques intéressantes de l'application transport optimal s'il existe dont on cite une : "Si le coût est la distance euclidienne alors on ne pourra jamais trouver deux trajectoires qui se coupent". Bien que ces questions soient anciennes il fallait attendre deux siècles pour avoir des réponses satisfaisantes. Notamment en 1941 lorsque [Kantorovitch](#) (Prix Nobel d'économie) a donné une nouvelle reformulation mathématique du problème du transport optimal plus souple et facile à manipuler en relaxant la contrainte de Monge.

Pour comprendre pourquoi la théorie n'a pas pu se développer avant Kantorovitch remarquant que le problème de Monge n'admet pas toujours de solution, en effet l'ensemble des contraintes peut être vide (par exemple si μ est une mesure de Dirac et ν est une mesure absolument continue ou la moyenne de deux Dirac distincts). En général l'ensemble des contraintes de Monge n'est pas convexe et même s'il est non vide on n'aura pas toujours l'unicité (voir la figure 2).

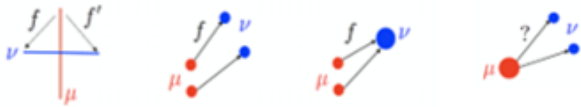


FIGURE 2 | Source : [Flamary and Courty]

1.2. Formulation de Kantorovich (1942)

Dans son étude au problème d'allocation des ressources, Kantorovitch a été amené à étudier un problème de transport optimal entre deux mesures discrètes. Mais dans ce cas, lors du transport des ressources un agent économique peut transférer sa richesse sur plus qu'un agent cible. Et c'est là où réside la différence fondamentale entre la formulation de Kantorovitch et celle de Monge. Dans la formulation de Kantorovitch on peut diviser la masse d'un point x de la distribution de départ de tel sorte qu'une partie de cette masse sera transporté à plusieurs points de la distribution cible. Par conséquent, on peut transporter la mesure μ vers ν de l'exemple 3 alors que ce n'était pas possible dans le cas de Monge.

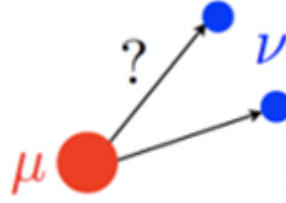


FIGURE 3 | Source : [Flamary and Courty]

Remarquons que dans ce nouveaux point de vue, le transport entre deux mesures ne pourra pas être décrit par une application de transport (car la masse présente en x peut se répartir dans plusieurs positions de l'espace). Mais plutôt par une notion plus générale celle du Couplage entre deux mesures.

1.2.1. Noyaux et couplages.

L'idée de permettre de casser la masse dans un point x peut être formalisé en associant à chaque point x de la distribution de départ une distribution de probabilité p_x qui modélise la distribution de masse transportée de x dans l'espace Y . l'application p qui associe à chaque $x \mapsto p_x$ est appelé un noyau sur X .

Etant donné une mesure source μ et un noyau p , alors on peut transporter μ par p en une mesure ν de tel sorte que :

$$\text{Pour tout borélien } \mathcal{B}; \nu(\mathcal{B}) = \int p_x(\mathcal{B}) d\mu(x)$$

Qu'on la note par : $\nu = p \star \mu$.

La masse transportée vers \mathcal{B} depuis μ est la somme des masses transportées vers \mathcal{B} depuis chaque point x de l'espace.

Notons que pour que ν soit définie, la définition précédente de noyau n'est pas suffisante. Pour cette raison on exige de plus que pour toute fonction f mesurable bornée que la fonction : $x \mapsto \int f(y) p_x(dy)$ est mesurable.

Enfin, la formulation de Kantorovitch consiste à remplacer la notion de transport de mesure dans celle de Monge par le noyau de transition p :

$$\inf_{\nu=p\star\mu} \int_{\Omega \times \Omega} c(x, y) dp_x(y) d\mu(x)$$

Or, cette formulation n'est pas souvent présentée dans la littérature et elle n'est qu'une voie équivalente et instructive à la formulation classique et historique de Kantorovitch. La formulation de Kantorovitch est la suivante :

$$(K) : \inf_{\pi \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) d\pi(x, y)$$

Avec $\Pi(\mu, \nu)$ désigne l'ensemble des mesure de probabilités sur $\Omega \times \Omega$ tel que sa première marginale est μ et sa deuxième marginale est ν .

Pour illustrer cette notion de couplage : remarquons que dans la figure 11 un couplage entre μ et ν est une densité sur l'espace produit représenté en orange. Déplacer μ à ν par ce couplage consiste à dépasser la masse de μ selon l'axe y de tel façon qu'on reconstruit la répartition orange. Ensuite déplacer la masse orange selon l'axe x de tel façon qu'on reconstruit ν .

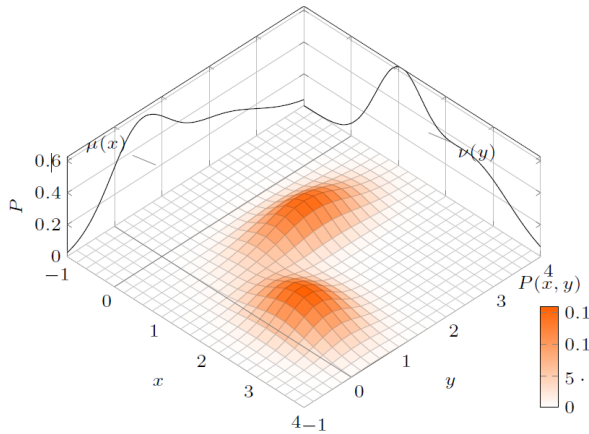


FIGURE 4 | Source : [Cuturi]

Pour voir le lien entre la formulation par noyaux et la formulation par couplages il est utile de voir μ et ν comme la loi de deux variables aléatoires.

Si X une variable aléatoire qui suit la loi μ et Y une variable aléatoire qui suit p_x conditionnellement à $X = x$, alors Y suit la loi $\nu = p \star \mu$. Ainsi le couple (X, Y) suit la loi π donné par : $\pi(A \times B) = \int_A p_x(B) d\mu(x)$. De plus π est un couplage entre μ et $\nu = p \star \mu$. Autrement dit à partir d'un noyau et une mesure source μ on peut obtenir un couplage entre μ et $\nu = p \star \mu$.

Réciproquement, à partir d'un couplage π entre μ et ν on peut trouver un noyau de tel sorte que $\nu = p \star \mu$ et $\pi(A \times B) = \int_A p_x(B) d\mu(x)$. De plus ce noyau est unique μ -presque partout (c'est exactement la conditionnelle de Y sachant X avec (X, Y) est un couple de variable aléatoire distribuer selon π).

Ainsi dans le cas où μ et ν sont soit à la fois discrètes ou continues on peut expliciter la loi de ce noyau en fonction des lois de μ et ν [Nathael Gozlan, 2018].

Finissons cette partie en citant deux couplages particuliers :

Le couplage indépendant Correspond à la loi jointe de X et Y indépendantes distribué selon μ et ν respectivement. En termes de noyau on a : $p_x = \nu$ pour tout x (tous les éléments de masse se repartissent selon ν indépendamment de l'endroit d'où ils partent).

Le couplage déterministe. C'est le cas où il existe une application T de transport de μ vers ν tel que le couplage est la loi du couple $(X, T(X))$ avec X suit la loi μ . En termes de noyau on a : $p_x = \delta_{T(x)}$.

1.2.2. Avantages théoriques.

Contrairement au problème de Monge, le problème de Kantorovitch est bien posé. Cela est dû au fait que l'ensemble des contraintes est non vide (couplage indépendant). Ainsi, le problème de Kantorovitch est symétrique par rapport à μ et ν .

De plus le problème de Kantorovitch est un problème d'optimisation linéaire en dimension infinie (dimension finie si les deux mesures sont discrètes) et donc on peut appliquer les méthodes de la théorie du calcul des variations à ce problème et en particulier à l'existence d'un minimum en montrant sous une condition de continuité où semi-continuité sur le coût et avec un argument de compacité qu'une séquence minimisante de Kantorovitch converge vers un plan optimal (Avec une notion propre de convergence dans l'espace des mesures de probabilité) [Nathael Gozlan, 2018].

Théorème 1 (Existence des plans optimaux). Soit $c : \Omega \times \Omega \rightarrow \mathbb{R}^+$ une fonction de coût semi-continue inférieurement. Si Ω est un espace polonais (séparable et complet) ; Alors, il existe un plan de transport optimal entre μ et ν . [Santambrogio, b]

Compte tenu de ce qui précède, le problème Kantorovitch est plus facile à traiter du point de vue théorique. Cela vient du fait que Kantorovitch a trop relaxé le problème de Monge. Notez également que Kantorovitch n'est pas une version plus générale que Monge (résoudre Kantorovitch (K) n'implique pas résoudre Monge (M)). Nous pouvons remarquer que la valeur minimale de Kantorovitch est toujours inférieure où égale à celle de Monge (avec la convention $\inf(M) = +\infty$ si l'ensemble des contraintes est vide). Par ailleurs, nous pouvons montrer que si un plan optimal de (K) est déterministe associé au transport optimal T alors $\inf(M) = \inf(K)$ et T est aussi un transport optimal de (M) (la réciproque reste vraie). Par conséquent, la résolution du problème de Monge revient à résoudre le problème de Kantorovitch et à voir si sa solution est déterministe [Santambrogio, b].

En complément de toutes ces considérations, qui montrent l'avantage de la formulation de Kantorovitch, nous pouvons introduire un problème dual de Kantorovitch, que l'on note (D). Cette dualité est sans saut (c'est-à-dire $\inf(K) = \sup(D)$) pour des hypothèses particulières sur le coût, ce qui a permis de caractériser des plans optimaux.

1.3. Problème dual

Une des grandes conséquences de la formulation de Kantorovitch est qu'on peut lui associer un problème dual (D). Or, notre problème est d'optimisation sous contraintes convexes. La forme dual s'écrit comme suit :

Définition 1 (Dualité de Kantorovich). *Soit la fonction coût $c : \Omega \times \Omega \rightarrow \mathbb{R}^+$, μ et ν deux densités de probabilités. La forme dual (D) est :*

$$(D) : \sup_{\phi(x)+\psi(y) \leq c(x,y)} \int \phi d\mu + \int \psi d\nu$$

Où $(\phi, \psi) \in L_1(\mu) \times L_1(\nu)$.

Définition 2. *Soit $\chi : \Omega \rightarrow \mathbb{R}$ une fonction, on définit la c -transformé concave de χ par :*

$$\chi^c(y) = \inf_{x \in \Omega} c(x, y) - \chi(x)$$

La formulation dual est intéressante car il est plus simple de manipuler des fonctions que manipuler des couplages et surtout d'un point de vue numérique où il est plus facile de stocker des fonctions dans la mémoire qu'à stocker des couplages.

Des questions naturelles se posent maintenant : Est-ce que le problème admet des solutions ? Est-ce qu'on peut les caractériser ? Et si oui, quelle relation avec les solutions optimales de (D) et le transport optimal de (K) ?

Commençons par l'existence, la première difficulté de cette formulation dual est que l'ensemble des contraintes n'est pas compact et que nous ne pouvons donc pas obtenir facilement un résultat d'existence comme nous l'avons vu avec (K). Pour surmonter cette difficulté, nous pouvons réécrire (D) sous une autre forme avec un ensemble de contraintes compact et une seule fonction c -concave au lieu de deux

$$\sup(D) = \sup_{\psi \in \Psi_c(\Omega)} \int \psi d\mu + \int \psi^c d\nu \quad (2)$$

Où $\Psi_c(\Omega)$ est l'ensemble des fonctions c -concaves.

Ainsi cette égalité a lieu (c'est-à-dire une égalité de dualité et de l'existence d'une solution optimale de D qu'on l'appelle un potentiel de Kantorovitch) si Ω est polonais et c uniformément continue. [Nathael Gozlan, 2018]

Une autre condition suffisante pour satisfaire ce résultat est Ω compact et c continue. [Santambrogio, a]

Notons que dans le cas très général $\inf(K)$ est inférieur à $\sup(D)$ et la différence entre ces deux quantités s'appelle le saut de dualité entre (K) et (D).

La nouvelle formulation ci-dessus avec des fonctions c -concave montre que la valeur du transport optimal $\inf K$ est convexe par rapport à μ et ν .

1.3.1. Potentiels de Kantorovitch.

Définition 3. *L'ensemble des fonctions ψ maximisant le problème 2 sont appelées les potentiels de Kantorovitch pour le transport de μ à ν .*

Maintenant, nous souhaitons caractériser les potentiels de Kantorovitch pour résoudre le problème primal. Cependant, c'est très difficile car nous ne savons pas comment le faire avec des coûts généraux. Mais nous pouvons trouver des résultats remarquables si nous nous intéressons à un coût proportionnel à une puissance de la distance (ce qui n'est déjà pas mal car de tels coûts couvrent la majorité des situations pratiques où l'on s'intéresse aux problèmes de transport).

Dans ce cas on peut montrer [Santambrogio, b] que sur un compact bornée les fonctions c -concaves sont : pD^{p-1} -lipschitzienne, avec p la puissance de la distance dans le coût et D le diamètre d'un tel domaine. De plus, lorsque $p = 1$ et $p = 2$, la fonction de coût est $c = (d(x, y)^p / p)$. On obtient les caractérisations suivantes :

$$\psi \in \Psi_1(\Omega) \iff \psi \text{ est } 1\text{-lipschitzienne}$$

$$\psi \in \Psi_2(\Omega) \implies x \mapsto \frac{x^2}{2} - \psi(x) \text{ est convexe} \quad (3)$$

On suppose que Ω est un ouvert compact de \mathbb{R}^d .

Théorème 2 (Cas d'un coût de la forme $h(x - y)$ où h est strictement convexe). *Si la mesure μ est à densité par rapport à la mesure de Lebesgue et la frontière de Ω est μ -négligeable, et il existe au moins un plan optimal qui couple μ et ν , alors :*

- Le plan optimal qui couple μ et ν est unique, et il est déterministe (On note T l'application transport associé à ce plan).
- Il existe un potentiel ψ de Kantorovitch dont le gradient est unique μ -presque partout.

- T et ψ sont reliés par la formule suivante μ -presque partout :

$$T(x) = x - (\nabla h^*) (\nabla \psi(x)) \quad (4)$$

Réciproquement, si une application peut s'écrire de la forme 4 μ -presque partout avec ψ une fonction c concave, alors cette application est un transport optimal du problème du Monge [Nathael Gozlan, 2018, Santambrogio, b].

Remarques. [Nathael Gozlan, 2018, Santambrogio, b]

- Si μ est positive. Alors, on peut prouver que le potentiel ψ est unique à une constante additive près.
- Si μ est à densité et T une application vérifiant 4 μ -presque partout avec ψ une fonction c -concave. Alors, T est un transport optimal entre μ et l'image de μ par T .
- Le résultat reste vrai si on suppose que ν est à densité au lieu de μ à condition de remplacer μ -presque partout par ν -presque partout.
- Si Ω est compact l'hypothèse d'existence d'un plan optimal sera automatiquement satisfaite.

En particulier, pour un coût de la forme $c(x, y) = h(x - y) = (d(x, y)^p/p)$ avec $p > 1$ la théorie 2 peut caractériser le transport optimal de Monge.

En utilisant la caractérisation des fonctions c -concaves, dans le cas où $p = 2$ et $\Omega = \mathbb{R}^d$. Alors, la caractérisation de T devient plus simple et $T = \nabla \psi$ où ψ est convexe (c'est le théorème de Brenier [Nathael Gozlan, 2018]).

Pour $p = 1$, le coût c ne peut pas être de la forme $h(x - y)$ avec h strictement convexe, et donc, on ne peut pas appliquer la théorie 2. Cependant, un théorème similaire peut être obtenu en ce qui concerne l'unicité du plan de transport optimal et son caractère déterministe, mais la caractérisation du transport optimal est plus compliquée et, en pratique, peu utile.

Deux formulations plus simples pour ce cas peuvent donc être obtenues en utilisant 3.

Proposition 1. L'équation 2 peut être reformulé comme :

$$\begin{aligned} \inf K &= \max_{\psi \text{ 1-lipschitzienne}} \int \psi d(\mu - \nu) \\ &= \text{Le minimum du problème du flou infimale de Beckmann [Santambrogio, a]} \end{aligned}$$

Quelques considération théorique. [Nathael Gozlan, 2018, Santambrogio, b] Finissons cette partie par quelques considération théorique :

- i Le théorème 2 reste vrai dans un ouvert de \mathbb{R}^d non compact à condition que ν soit à support compact et que de plus μ est à densité dont la frontière de l'intérieur de son support est négligeable et que la valeur $\inf(K)$ est finie.
- ii On peut remplacer la condition de μ est à densité par n'importe quelle condition générale qui assure que l'ensemble des points où un potentiel de Kantorovitch est non différentiable est négligeable.
- iii Notons qu'on aura encore l'unicité du plan optimal dans le théorème 2 sans que la frontière de X soit négligeable (cependant il ne sera pas forcément déterministe).
- iv Pour le cas d'un coût proportionnelle à la puissance d'une distance (non euclidienne), le théorème 2 reste vrai en supposant seulement que $\inf(K)$ est finie.
- v Un autre cas où le théorème de Bernier reste vrai est celui où deux variables aléatoires de lois respectives μ et ν admettent un moment d'ordre 2 et que μ est sans masse sur tout surface de dimension $d - 1$ de classe C^2 .
- vi On peut obtenir une version plus générale du théorème 2 avec des coûts plus généraux réguliers. Mais pour une introduction à la théorie du transport optimal et leurs applications le théorème 2 est largement suffisante.

1.4. Quelques cas intéressants

Dans son cadre le plus générale, le problème du transport optimal n'est pas facile à résoudre. Cependant, il existe deux cas très particuliers où on sait trouver d'une manière explicite ou calculer directement le transport optimal : c'est le cas de la dimension 1 et le cas du transport entre deux mesures discrètes. En effet, ces deux cas sont très intéressants en pratiques et permettent d'englober une panoplie d'application très diversifié.

1.4.1. Le problème du transport optimal discret

Dans le cas de deux mesures discrètes $\mu = \sum_{1 \leq i \leq m} p_i \delta_{x_i}$ et $\nu = \sum_{1 \leq j \leq n} q_j \delta_{y_j}$ dans le problème (K), un couplage Π sera de la forme :

$$\Pi = \sum_{i,j} \Pi_{ij} \delta_{\{x_i\} \times \{y_j\}} \quad \text{avec} \quad \begin{cases} \sum_j \Pi_{ij} = p_i \\ \sum_i \Pi_{ij} = q_j \\ \Pi_{ij} \geq 0 \end{cases}$$

Ainsi, le coût du transport sera entièrement déterminé par la matrice $C = (c_{ij} = c(x_i, y_j))$. D'où, le problème (K) sera dans ce cas un problème d'optimisation li-

néaire en dimension finie :

$$(P) : \begin{cases} \inf \langle c, \Pi \rangle = \sum_{i,j} c_{ij} \Pi_{ij} \\ \Pi \in \mathbb{R}^{m \times n} \\ \Pi \cdot \mathbb{1}_m = p \\ \Pi \cdot \mathbb{1}_n = q \\ \Pi \geq 0 \end{cases}$$

et son problème dual associé s'écrit :

$$(D) : \begin{cases} \inf \langle p, u \rangle + \langle q, v \rangle \\ u \in \mathbb{R}^m \\ v \in \mathbb{R}^n \\ u_i + v_j \leq c_{ij} \end{cases}$$

Vue que l'ensemble des contraintes est non vide, et bornée (couplage indépendant). Le problème (P) admet toujours une solution. plus généralement, la théorie de programmation linéaire nous annonce que :

- i (P) et (D) admettent au moins une solution et $\max D = \inf P$.
- ii Si Π est solution de (P) et (u, v) solution de (D), alors,

$$\Pi_{ij} > 0 \Rightarrow u_i + v_j = c_{ij}$$

- iii Si (u, v) solution de (D), alors :

$$\begin{cases} u_i = \inf_j (v_j - c_{ij}) \\ v_j = \inf_i (u_i - c_{ij}) \end{cases}$$

Nous verrons à la fin de cet article que ces résultats peuvent servir à quelques applications en économie.

Dans le cas où $n = m$ et $p_i = q_j = 1/n$, et grâce à un lemme . les solutions du problème (P) sont déterministe et résout le problème du Monge discret suivant (qui n'est rien qu'un problème d'assignement).

$$\inf_{\sigma \in \delta_n} \sum_i c_{i, \sigma(i)}$$

Enfin, le problème du transport optimal discret se ramène soit à un problème d'optimisation linéaire, où sous forme standard d'assignement. De tels problèmes sont beaucoup étudiés dans la littérature et peuvent être résolus par plusieurs algorithmes (simplexe, simplexe révisé, points intérieurs, flot minimal, hongrois, ...etc).

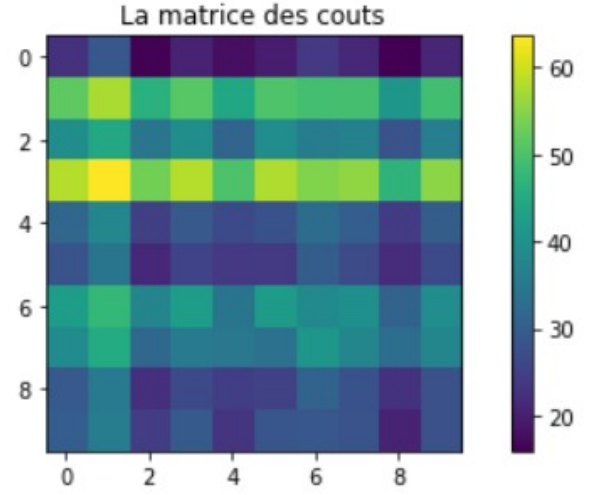


FIGURE 5 | Le coût du transport optimal

la matrice du couplage optimal est:

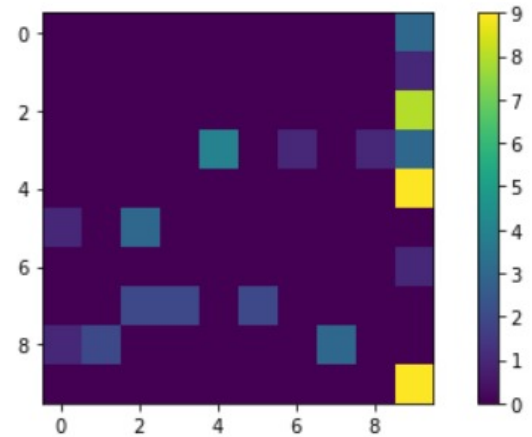


FIGURE 6 | Le coût du transport optimal

le cout du transport optimal est: 1910.0637372799292

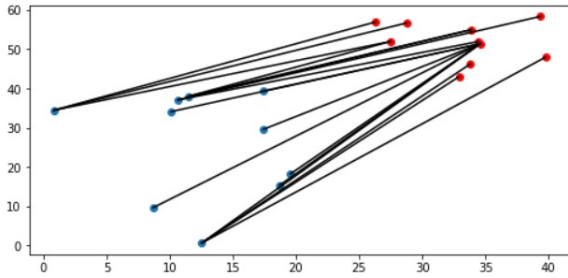


FIGURE 7 | Simulation du couplage optimal entre deux distributions aléatoires dans le plans. Source : [\[Our GitHub\]](#)

1.4.2. Le transport optimal sur la droite réelle

Dans le cas des mesures définies sur la droite réelle, la notion "d'ordre" nous permet à résoudre le problème du transport optimal d'une manière explicite.

Théorème 3. Notons par F^{-1} l'inverse généralisé de la fonction de répartition d'une mesure.

Si le coût est de la forme $h(x - y)$ où h est strictement convexe et $\inf K < \infty$, alors, il existe un unique transport optimal qui est monotone croissant (décroissant si h strictement concave).

De plus, si μ est sans atomes alors le plan monotone est déterministe et :

$$T = F_{\mu}^{-1} \circ F_{\nu}$$

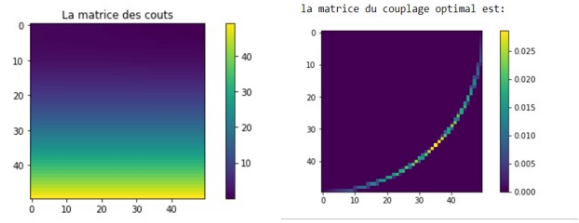
$$\inf K = \int_0^1 h \left(F_{\mu}^{-1}(t) - F_{\nu}^{-1}(t) \right) dt$$

Remarques 1. Si h n'est ni strictement convexe, ni strictement concave. Le résultats d'existence d'un plan monotone optimal reste vrai mais sans unicité en général.

On peut généraliser ces résultats pour des coûts qui vérifie des conditions de type :

$$c(x, y') + c(x', y) \leq c(x, y) + c(x', y')$$

Dans le cas discret, résoudre le problème du transport optimal en dimension 1 revient à faire un tri (coût de calcul $O(N \log N)$).



le cout du transport optimal est: 25.002468968181446

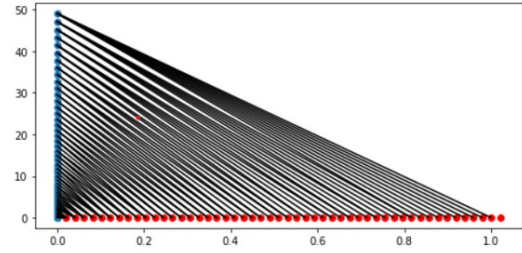


FIGURE 8 | Simulation d'un transport monotone dans le cas 1D. Source : [\[Our GitHub\]](#).

2. Applications du problème du transport optimal

Dans la littérature le transport optimal est utilisé comme un outil pour un large spectre de problème dans plusieurs branches de mathématiques que ça soit pure ou appliqués : la géométrie, les mathématiques de la physiques, la recherche opérationnelle, l'économie, la finance quantitative, le traitement d'images et des signaux et dans plusieurs problématiques en science des données et apprentissage automatique. En effet, la connexion entre la théorie du transport optimal et ces applications polydisciplinaires peut s'expliquer par le fait que le formalisme de cette théorie est lié directement aux notions géométriques et des déformations physiques usuelles qui sont intimement liée au principe du moindre action et à la conservation de masse. Au-delà, le problème du transport optimal permet de définir une distance sur l'espace des mesures de probabilités qui est plus naturel pour comparer et moyenner des distributions de probabilités tout en gardant la bonne intuition géométrique de ressemblance entre deux distributions dans un espace géométrique, chose qu'on ne le savait pas faire avec des métriques usuels (L^p par exemple).

Dans cette partie, nous introduisons la distance de Wasserstein en illustrant ses avantages sur des situations pratiques. Nous mettons en évidence les limites de la formulation classique du transport optimal en sciences des données et on présentera le concept de la régularisation entropique à travers le problème de moyennisation des données. Enfin, Nous illustrons comment le transport optimal permettra à modéliser des situations d'équilibres et de stabilités en économie.

2.1. Distances de Wasserstein

Dans le cas où Ω est un espace métrique compact et pour tout $p \geq 1$, on définit :

$$W_p(\mu, \nu) = (\inf(K))^{1/p} \text{ avec } c(x, y) = |x - y|^p$$

Où par dualité :

$$\frac{1}{p} W_p^p(\mu, \nu) = \sup_{\psi \in \Psi_p(\Omega)} \int_{\Omega} \psi d\nu + \int_{\Omega} \psi^c d\mu$$

W_p est une distance sur l'espace $\mathcal{P}(\Omega)$ et la topologie induite par cette distance correspond à la topologie de convergence faible des mesures de probabilités [Santambrogio, a].

Exemple illustratif. Dans le cas des mesures de probabilités gaussiennes, la distance euclidienne entre deux gaussiennes de même variance est constante dès que leurs supports deviennent disjoints, alors que la distance de Wasserstein est strictement monotone par rapport à la moyenne de la gaussienne cible. Autrement dit, la distance de Wasserstein capture l'intuition du mouvement collective de la masse contrairement aux distances usuels (L^p). Nous illustrons ce phénomène avec la simulation suivante :

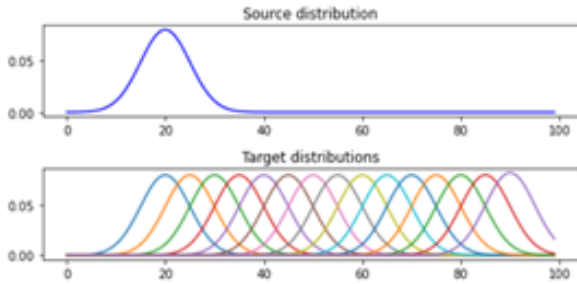


FIGURE 9 | Source : [Our GitHub].

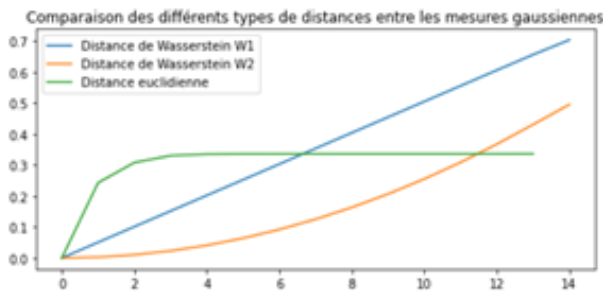


FIGURE 10 | Comparaison des différents types de distances entre les mesures gaussiennes. Source : [Our GitHub].

2.2. Le barycentre de Wasserstein et moyennisations des données empiriques

L'agrégation efficace de données provenant de différentes sources est un problème difficile, en particulier lorsque les échantillons de chaque source sont distribués différemment. L'une des façons de fusionner des distributions de probabilités est de le faire par le biais d'un transport optimal. La méthode du barycentre de Wasserstein [Gramfort et al.] est une distribution unique qui résume une collection de mesures d'entrée tout en respectant leur géométrie.

A travers cette application d'IRM, qui se base sur les idées introduites par Kantorovich [?] afin de proposer un nouvel algorithme capable de calculer efficacement la moyenne de données non normalisées définies sur des domaines discrets arbitraires en utilisant des métriques de transport.

Elle montre comment les moyennes de Kantorovich peuvent être liées aux barycentres de Wasserstein [Agueh and Carlier] afin de tirer parti de l'approche du lissage entropique [Cuturi and Doucet] afin de profiter de l'approche du lissage entropique. Elle conduit à un problème d'optimisation convexe lisse et à un algorithme présentant de fortes garanties de convergence.

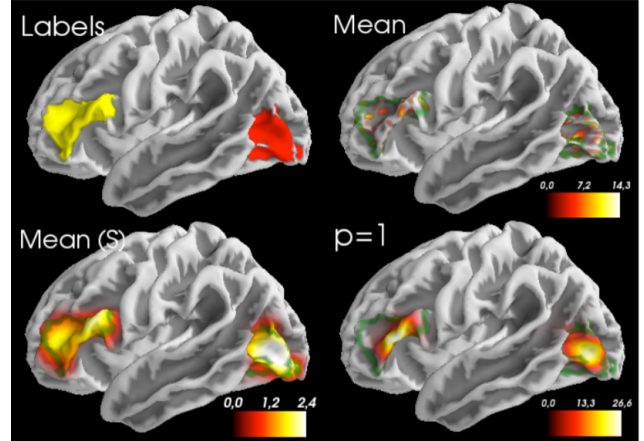


FIGURE 11 | Résultats de la simulation avec des signaux aléatoires générés dans un groupe de 100 sujets.

Cette application illustre la diversité de cet outil et son comportement empirique sur des données de neuroimagerie fonctionnelle, des estimations de sources en IRM fonctionnelle et en magnétoencéphalographie (MEG), définies sur des grilles de voxels et des triangulations de la surface corticale pliée.

Les données étant volumineuses et d'une grande variabilité, il est difficile de procéder directement à

leur lissage. Cependant, en utilisant cette méthode, nous aurons la possibilité de lisser les données que ce soit leurs variabilités et leurs complexités.

Les contributions de cette méthode sont doubles. Premièrement, en considérant des mesures non-normalisées, particulièrement pertinentes pour les données d'imagerie médicale [Thirion et al.], nous étendons l'état de l'art actuel dans l'estimation du barycentre en utilisant des métriques de transport.

Deuxièmement, à la suite de contributions récentes sur le transport optimal discret, cette méthode propose une version lissée du problème de transport qui conduit à un algorithme d'optimisation efficace. Alors que de nombreuses contributions sur le transport optimal ne travaillent que dans une ou deux dimensions sur une grille régulière, cette approche peut s'adapter à la géométrie complexe du cerveau (grilles et surfaces irrégulières). Seule la définition d'une métrique est nécessaire. L'algorithme proposé implique des opérations simples qui particulièrement adaptées au matériel GPU moderne et qui permet de calculer les barycentres sur des données cérébrales complètes en quelques minutes. Cette méthode a montré la capacité à mettre clairement en évidence les foyers d'activation tout en évitant le besoin de lisser les données.

2.3. Application en économie

Bien qu'elle a été développée par Kantorovitch lorsqu'il a été confronté à des problèmes économiques, la théorie du transport optimal est presque rarement citée d'une manière explicite dans la littérature économique. C'est seulement dans la dernière décennie que Galichon a pu mettre en évidence la connexion directe entre le formalisme du transport optimal et des problèmes économiques fondamentales (les problèmes d'allocations, d'équilibre du marché, d'inversion de demande, et les quantiles multivariés ...etc).

En microéconomie, si on suppose que chaque agent économique est rationnel : c'est-à-dire qu'il cherche à maximiser son utilité, on peut caractériser le prix d'équilibre à l'aide d'un problème d'optimisation (Minimisation des surplus des consommateurs et des producteurs, généralement non linéaire) en connaissant les utilités de chaque agent économique.

Cependant, si on décide de voir les agents économiques comme deux distributions de probabilités discrètes on peut associer à ce marché un couplage qui maximise une utilité globale en connaissant les surplus $\Phi(x, y)$ engendré par la consommation d'un agent x d'une unité de production y . Et cette fois ci on affronte un problème linéaire.

A première vue on pourrait dire que cette modéli-

sation cache la notion d'utilité et de rationalité d'un agent individuel. Nous illustrons à travers l'exemple suivant comment la théorie du transport pourra servir à modéliser ces notions seulement à l'aide d'un problème d'optimisation linéaire, ainsi qu'à caractériser des prix à l'équilibre.

Le problème d'affectation et du prix d'équilibre dans un marché. Supposons que nous avons un marché constitué par un ensemble Y de n producteurs d'un bien et d'un ensemble X de m consommateurs prêt à consommer ce bien et notons par q_y la quantité produite par un producteur y et p_x la quantité demandée par un consommateur x . On suppose que la demande est égale à l'offre ($\sum p_x = \sum q_y$). Si on note $\Phi(x, y)$ le surplus engendré entre les agents économiques si x achète une unité du bien du y . Le problème d'affectation consiste à trouver les quantités $\Pi(x, y)$ qui représente quantité du bien acheté du bien produit par y de la part des agents x de tel façon que le surplus total sera maximal. Formellement ce problème s'écrit comme un problème du transport optimal, c'est-à-dire :

$$\inf \sum_{xy} \pi_{xy} \Phi_{xy} \quad \text{avec} \begin{cases} \sum_y \pi_{xy} = p_x \\ \sum_x \pi_{xy} = q_y \end{cases}$$

Cette notion d'affectation peut être utilisée à plusieurs situations économiques par exemple X des travailleurs et Y des entreprises qui offrent des positions d'emplois pour ces travailleurs.

Cette fois-ci le problème d'optimisation est linéaire et qui nous permettra de plus définir une notion d'équilibre à travers sa formulation dual. En effet les propriétés i., ii. et iii. en 1.4.1, nous donne l'intuition d'interpréter les solutions du problème dual associé comme des utilités marginales associés aux agents économiques du marché qui cherche à les maximiser. En effet, la condition $\mu_i + \nu_j$ supérieur $\Phi(x, y)$ peut s'interpréter par une situation de stabilité car si $\mu_i + \nu_j < \Phi(x, y)$ alors les deux agents x_i et y_j ne sont pas en couple d'après ii. en 1.4.1 et ont intérêt à se mettre en couple pour générer un surplus $\mu'_i + \nu'_j = \Phi(x, y)$ plus supérieur à celui qu'avant.

Hypothèses et notations.

- Supposons que $\Phi(x, y) = \alpha_{xy} + \beta_{xy}$ avec α_{xy} représente l'utilité de x d'acheter le bien de y et β_{xy} représente l'utilité de y de vendre son bien à x .
- Notons w_{xy} le prix que paye x à y s'il achète une unité de produit.

Théorème 4 ([Galichon]). *Si on admet que les potentiels de Kantorovitch associé au problème dual du (P) représente les utilités des agents et que pour le système des prix w_{xy} à l'équilibre est donné par la maximisation des utilités des agents, c'est-à-dire que :*

$$\begin{aligned}\mu_x &= \max_y \alpha_{xy} + w_{xy} \\ \nu_y &= \max_x \beta_{xy} - w_{xy}\end{aligned}$$

Alors le système des prix w_{xy} peut être caractérisé par les solutions (μ, ν) du problème dual du (D).

Si $\Pi_{xy} > 0$, alors :

$$\begin{aligned}w_{xy} &= \beta_{xy} - \nu_y \\ &= \mu_x - \alpha_{xy}\end{aligned}$$

3. Conclusion

Le problème du transport optimal même s'il est ancien n'a cessé d'intéresser les mathématiciens et les scientifiques d'autres disciplines. En effet, cette belle théorie donne un cadre approprié pour capturer des intuitions géométriques complexes et naturelles. Ainsi que sa souplesse en termes d'hypothèses sur les couts et la nature des espaces ou on définit ce transport. Cette souplesse et cette image géométrique que le transport optimal peut capturer dans plusieurs situations complexes ont permis à cette théorie d'être une boîte à outils de nombreux problèmes polydisciplinaires, surtout après l'arrivée des méthodes numériques puissantes pour le résoudre d'une manière approchée qui ont donné à cette théorie des nouvelles applications pratiques en sciences des données. Dans ce modeste article, nous avons essayé d'introduire et synthétiser dans des cadres généraux le plus possible les bases théoriques de cette théorie, ainsi qu'à montré par quelques exemples comment cette dernière peut servir à des problèmes déconnectés les uns aux autres à première vue. Finalement, il serait judicieux d'examiner le cas continu et semi-continu afin de pouvoir comprendre le lien entre cette théorie et d'autres domaines telle que la mécanique des fluides et la théorie des jeux.

Remerciements

Nous tenons à remercier notre encadrant de projet M. Adil Ahidar, enseignant-chercheur en mathématique, pour tout le soutien, le suivi, l'aide, l'orientation et la guidance sans lesquels nous n'aurions pas découvert et comprendre les bases de cette belle théorie.

Références

- M. Agueh and G. Carlier. Barycenters in the wasserstein space. 43(2) :904–924. doi: 10.1137/100805741. URL <https://hal.archives-ouvertes.fr/hal-00637399>. Publisher : Society for Industrial and Applied Mathematics.
- L. Ambrosio and N. Gigli. *A User's Guide to Optimal Transport*, volume 2062, pages 1–155. Springer Berlin Heidelberg. ISBN 978-3-642-32159-7 978-3-642-32160-3. doi: 10.1007/978-3-642-32160-3_1. URL http://link.springer.com/10.1007/978-3-642-32160-3_1. Series Title : Lecture Notes in Mathematics.
- M. Cuturi. Fast computation of wasserstein distances and applications to parameter estimation. page 68.
- M. Cuturi and A. Doucet. Fast computation of wasserstein barycenters. URL <http://arxiv.org/abs/1310.4375>.
- R. Flamary and N. Courty. Optimal transport for machine learning - introduction to optimal transport. page 72.
- A. Galichon. *Optimal transport methods in economics*. Princeton University Press. ISBN 978-0-691-17276-7. OCLC : ocn944469171.
- A. Gramfort, G. Peyré, and M. Cuturi. Fast optimal transport averaging of neuroimaging data. URL <http://arxiv.org/abs/1503.08596>.
- P. Henry-Labordère. *Model-Free Hedging : A Martingale Optimal Transport Viewpoint*. Chapman and Hall/CRC, 1 edition. ISBN 978-1-315-16174-7. doi: 10.1201/9781315161747. URL <https://www.taylorfrancis.com/books/97813151666237>.
- L. Kantorovich and G. S. Rubinstein. On a space of totally additive functions. *Vestnik Leningrad. Univ*, 13 :52–59, 1958.
- L. Kantorovitch. On the translocation of masses. 5(1) :1–4. URL <http://www.jstor.org/stable/2626967>.
- P.-A. Z. Nathael Gozlan, Paul-Marie Samson. Notes de cours sur le transport optimal. page 115, 2018.
- N. Papadakis. Optimal transport for image processing. page 112.
- I. Redko. Transfer learning and optimal transport. page 128.
- F. Santambrogio. Introduction to optimal transport theory. a. URL <http://arxiv.org/abs/1009.3856>.
- F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Springer International Publishing, b. ISBN 978-3-319-20827-5 978-3-319-20828-2. doi: 10.1007/978-3-319-20828-2. URL <http://link.springer.com/10.1007/978-3-319-20828-2>.
- B. Thirion, P. Pinel, S. Mériaux, A. Roche, S. Dehaene, and J.-B. Poline. Analysis of a large fMRI cohort : Statistical and methodological issues for group analyses. 35(1) :105–120. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2006.11.054.