

# Théorie du transport optimal

Racine Florian

January 17, 2023

# Table des matières

Chapter 1	Cours	Page 2
1.1	Introduction Formulation du problème — 2	2
1.2	Modélisation	3
1.3	La formulation du problème de transfert optimal de Monge	6
1.4	La dualité de Kantorovitch La théorie — 8 • Application(s) — 10	8
1.5	La distance de Wasserstein	11
1.6	Estimation numérique du transfert optimal L'algorithme de Sinkhorn — 14 • Formulation dynamique — 16	14
Chapter 2	TD	Page 18
2.1	TD1	18

# Chapter 1

## Cours

### 1.1 Introduction

#### 1.1.1 Formulation du problème

##### Question 1

Quelle est la façon optimal de transporter un tas de sable dans un trou ?

##### Question 2

Comment construire un chateau de sable d'une forme données à partir d'un tas de sable ?

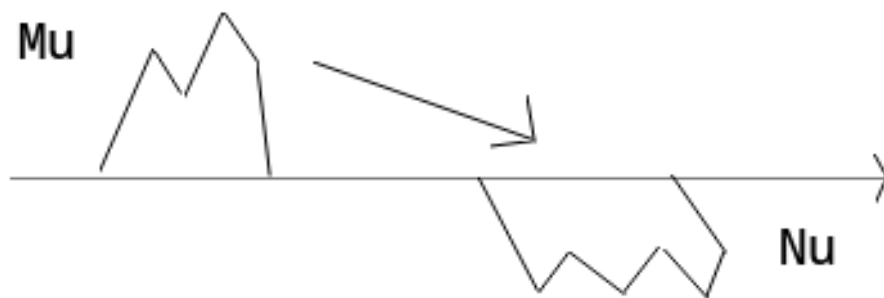


Figure 1.1: Transporter un tas de sable dans un trou

##### Note:-

Avec le même nombre de grain de sable et la même masse.

## 1.2 Modélisation

$\nu \in \mathcal{P}(\mathbb{R}) ; \mu \in \mathcal{P}(\mathbb{R})$

### Definition 1.2.1

$\forall A \in \mathcal{P}(\mathbb{R}), \mu[A]$  décrit quelle quantité de sable est dans  $A$ .

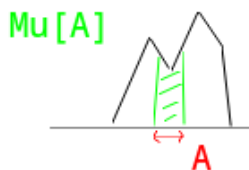


Figure 1.2:  $\mu[A]$

### Definition 1.2.2

Cout infinitésimal :

$$C : \begin{cases} \mathbb{R} \times \mathbb{R} & \longrightarrow \mathbb{R} \\ (x, y) & \longmapsto C(x, y) \end{cases}$$

Cout de transporter un grain de sable de  $x$  vers  $y$ .

### Question 3

Comment transporter un tas de sable avec un cout global minimal ?

### Definition 1.2.3: Tribu

Soit  $\Omega$  un espace. On note  $\mathcal{P}(\Omega)$  l'ensemble des parties de  $\Omega$ . Un ensemble d'ensembles  $\mathcal{A} \subset \mathcal{P}(\Omega)$  est appelé **tribu** s'il vérifie les trois propriétés suivantes:

- $\emptyset \in \mathcal{A}$  (contient l'ensemble vide)
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$  (stable par complémentaire)
- $(A_n)_{n \geq 1} \subset \mathcal{A} \Rightarrow \cup_{n \geq 1} A_n \in \mathcal{A}$  (stable par union dénombrable)

### Definition 1.2.4: Mesure

Soit,  $(\Omega, \mathcal{A})$  un ensemble mesurable. On appelle mesure positive sur  $(\Omega, \mathcal{A})$  une application  $\mu$  de  $\mathcal{A}$  à valeurs dans  $[0; +\infty]$  telle que:

- $\mu(\emptyset) = 0$  (la mesure de l'ensemble nul est égale à 0)
- $\forall A, B \in \mathcal{A} | A \cap B = \emptyset, \mu(A \cup B) = \mu(A) + \mu(B)$  (la mesure de deux ensembles disjoints est la somme des mesures)
- $\forall$  Suite croissante  $(A_n)_{n \geq 1} \subset \mathcal{A}, \lim_{n \rightarrow \infty} \mu(A_n) = \mu(\cup_{n \geq 1} A_n)$

#### Note:-

Le triplet  $(\Omega, \mathcal{A}, \mu)$  est appelé espace mesurable.

### Definition 1.2.5: Plan de transport

Un plan de transport entre les mesures  $\mu$  et  $\nu$  est une mesure de probabilité :  
 $\pi \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$  à pour marginale  $\mu$  et  $\nu$ .

#### Note:-

$\pi \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$  à pour marginal  $\mu$  et  $\nu$

$$\Leftrightarrow \forall A, B \text{ ensemble mesurable avec } A \subset \mathbb{R} \text{ et } B \subset \mathbb{R} \begin{cases} \pi[A \times \mathbb{R}] = \mu[A] \\ \pi[\mathbb{R} \times B] = \nu[B] \end{cases}$$

$$\Leftrightarrow \forall \varphi \in C^0(\mathbb{R}), \Psi \in C^0(\mathbb{R}) : \int_{\mathbb{R} \times \mathbb{R}} \varphi(x) + \Psi(y) d\pi(x, y) = \int \varphi(x) d\mu(x) + \int \Psi(y) d\nu(y)$$

#### Note:-

On notera,  $\Pi(\mu, \nu) = \{ \pi \in \mathcal{P}(\mathbb{R} \times \mathbb{R}) | \pi \text{ a pour marginal, } \mu, \nu \}$

On remarquera que,  $\Pi(\mu, \nu) \neq \emptyset$

$$I[\pi] = \int_{\mathbb{R}^2} C(x, y) d\pi(x, y) \text{ Le cout total associé au plan de transport optimal.}$$

On cherche,  $\tau_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} (I[\pi])$

### Definition 1.2.6: Plan de transfert optimal

S'il existe  $\pi_0 \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$  tel que  $I[\pi_0] = \tau_c(\mu, \nu)$  (plan de transport qui minimise le cout total)  
 $\pi_0$  est appelé un **plan de transfert optimal**

#### Exemple 1.2.1 (Exemple trivial (Kantorovitch))

$$a < b$$

$$c < d$$

$$C(x, y) = |x - y|^2$$

$$\mu = \frac{1}{2}(\delta_a + \delta_b)$$

$$\nu = \frac{1}{2}(\delta_c + \delta_d)$$

#### Question 4

Trouver :  $\pi(\mu, \nu) = ?$

**Solution:**  $\pi_\alpha = \frac{1}{2}(\alpha\delta_{(a,c)} + (1-\alpha)\delta_{(a,d)} + (1-\alpha)\delta_{(b,c)} + \alpha\delta_{(b,d)})$   $\Pi(\mu, \nu) = \{\pi_\alpha | \alpha \in [0, 1]\}$

#### Question 5

Calculer :  $I[\pi] = ?, \forall \pi \in \Pi(\mu, \nu)$

**Solution:**  $I[\pi_\alpha] = \int_{\mathbb{R}^2} C(x, y) d\pi_\alpha(x, y)$

$$I[\pi_\alpha] = \frac{1}{2}(\alpha C(a, c) + (1-\alpha)C(a, d) + (1-\alpha)C(b, c) + \alpha C(b, d))$$

$$I[\pi_\alpha] = \frac{1}{2}(a^2 + b^2 + c^2 + d^2) - \alpha(ac + bd) - (1-\alpha)(ad + cb)$$

#### Question 6

Trouver :  $\tau_c(\mu, \nu) = ?$

**Solution:**  $P(\alpha) = \frac{\partial I[\pi_\alpha]}{\partial \alpha} \implies P(\alpha) = -ac - bd + ad + cb; \implies P(\alpha) = (d-c)(a-b) < 0$   
Donc,  $I[\pi_\alpha]$  atteint son min en  $\alpha = 1$

$\pi_0 = \pi_{\alpha=1} = \tau_c(\mu, \nu)$   
 Donc,  $a \rightarrow c$  et,  $b \rightarrow d$

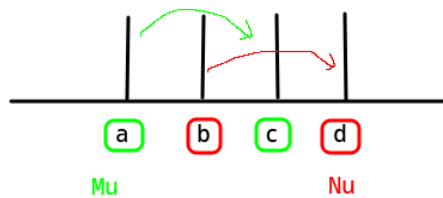


Figure 1.3: Solution de l'exemple de Kotorovitch

## 1.3 La formulation du problème de transfert optimal de Monge

### Note:-

On autorise pas le fait de couper les masses. A chaque  $x$  est associé une unique  $y$ .  
On dit que  $T$  envoie  $\mu$  sur  $\nu$  et on note :  $T\#\mu = \nu$

### Proposition 1.3.1

$\forall A \subset \mathbb{R}$  partie mesurable :  $\nu(A) = \mu(T^{-1}(A))$

$\Leftrightarrow$

### Proposition 1.3.2

$\forall \varphi$  continue :  $\int_{\mathbb{R}} \varphi(y) d\nu(y) = \int_{\mathbb{R}} (\varphi \circ T)(x) d\mu(x)$   
 $\tau_c^M(\mu, \nu) = \inf_{T: T\#\mu = \nu} I[T]$   
 $I(T) = \int_{\mathbb{R}} C(x, T(x)) d\mu(x)$

### Note:-

Solution de cout optimal d'après Kantorovitch  $\leq$  Solution de cout optimal d'après Monge  
Dans le première exemple ils coïncident.

### Note:-

Kantorovitch définit un problème linéaire en  $\Pi$ .  
Monge définit un problème non linéaire en  $T$ .

### Note:-

Problème de Kantorovitch admet toujours une solution  $\Pi_0$ .  
Problème de Monge n'admet pas toujours de solution n'y même d'application qui envoie  $\mu$  sur  $\nu$ .

### Exemple 1.3.1

$$\begin{cases} \mu \in \mathcal{P}(\mathbb{R}) \\ \nu = \delta_a \end{cases}$$

**Kantorovitch** :  $\Pi(\mu, \nu) = \{\mu \otimes \delta_a\}$

**Monge** : Quelles sont les  $T$  tel que  $T\#\mu = \nu$  ?

Il en existe une seule:

$$\forall x | T : \begin{cases} x & \longrightarrow & a \\ \mathbb{R} & \longmapsto & \mathbb{R} \end{cases}$$

$$\tau_c^M(\mu, \nu) = \tau_c(\mu, \nu)$$

D'une part :

$$\tau_c^M(\mu, \nu) = \int_{\mathbb{R}} C(x, T(x)) d\mu(x)$$

$$\tau_c^M(\mu, \nu) = \int_{\mathbb{R}} C(x, a) d\mu(x)$$

D'autre part :

$$\tau_c(\mu, \nu) = \int_{\mathbb{R}} C(x, y) d\Pi(x, y)$$

$$\tau_c(\mu, \nu) = \int_{\mathbb{R}} C(x, y) d(\mu \otimes \delta_a)(x, y)$$

$$\tau_c(\mu, \nu) = \int_{\mathbb{R}} C(x, y) d\mu(x) d\delta_a$$

$$\tau_c(\mu, \nu) = \int_{\mathbb{R}} C(x, a) d\mu(x)$$

**Note:-**

Le problème revient à empiler des grains de sables à partir d'un tas de sable. Dans ce cas les solutions de Kantorovitch et de Monge coïncident.

**Exemple 1.3.2**

$$\begin{cases} \mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \\ \nu = \frac{1}{n} \sum_{i=1}^n \delta_{y_i} \end{cases}$$

Les plans de transports  $\Pi$  entre  $\mu$  et  $\nu$  peuvent être représenté par des matrices bistochastiques de tailles  $n$ .

**Definition 1.3.1: Matrice bistochastique**

$$\begin{aligned} 0 &\leq \Pi_{i,j} \leq 1 \\ \sum_{i=1}^n \Pi_{i,j} &= 1 \text{ (la somme des éléments de chaque ligne vaut 1)} \\ \sum_{j=1}^n \Pi_{i,j} &= 1 \text{ (la somme des éléments de chaque colonne vaut 1)} \end{aligned}$$

On note  $\mathcal{B}_n$  l'ensemble des matrices bistochastiques.

Soit  $\Pi \in \mathcal{B}_n : I[\Pi] = \frac{1}{n} \sum_{i=1}^n C(x_i, y_i) \Pi_{i,j}$

$$\tau_c(\mu, \nu) = \inf_{\Pi \in \mathcal{B}_n} \left\{ \frac{1}{n} \sum_{i=1}^n \Pi_{i,j} C(x_i, y_i) \right\}$$

Il s'agit d'un problème linéaire de minimisation sur un ensemble convexe.

**Proposition 1.3.3 Ensemble convexe**

$\mathcal{B}_n$  est convexe  $\Leftrightarrow A, B \in \mathcal{B}_n$  alors  $\forall \theta \in [0, 1] \theta A + (1 - \theta)B \in \mathcal{B}_n$

**Definition 1.3.2: Points extrémaux**

L'ensemble des points extrémaux de  $E$  convexe est l'ensemble des  $e \in E$  tel que :

si  $e = \theta e_1 + (1 - \theta)e_2$  avec  $\theta \in [0, 1], e_1 \in E, e_2 \in E$

Alors  $\theta = 0$  ou  $\theta = 1$

**Théoreme 1.3.1 Théorème de Choquet**

$F$  est linéaire sur un domaine  $K$  convexe et compact, alors  $F$  admet au moins un minimum. Parmi les minimums de  $F$  au moins l'un d'eux est un extréma de  $K$ .

**Théoreme 1.3.2 Théorème de Birkhoff**

$\mathcal{B}_n$  est convexe et compact.

$\mathcal{B}_n$  admet  $n$  points extrémaux qui sont les matrices de permutations

Ainsi, le min pour le problème de Kantorovitch est atteint pour  $\begin{cases} \Pi_{i,j} = 1 & \text{si } ij = \sigma(i) \\ \Pi_{i,j} = 0 & \text{sinon} \end{cases}$



## 1.4 La dualité de Kantorovitch

### 1.4.1 La théorie

**Note:-**

Introduction :

Le problème primal de Kantorovich est un problème d'optimisation qui consiste à minimiser le coût de transport, tandis que le problème dual de Kantorovich est un problème d'optimisation qui consiste à maximiser la valeur de la fonction objectif du problème primal en utilisant les mêmes contraintes. Les deux problèmes sont liés par la théorie de dualité de Kantorovitch.

#### **Théoreme 1.4.1** Dualité de Kantorovitch

$\mu \in \mathcal{P}(\mathbb{R}^n); \nu \in \mathcal{P}(\mathbb{R}^n)$

C semi continue inférieurement ( par exemple C continue)

**Note:-**

$$\text{Pour } \pi \in \Pi(\mu, \nu), I[\pi] = \int C(x, y) d\pi(x, y)$$

Soit  $(\varphi, \psi) \in \phi_c$

$$J(\varphi, \psi) = \int_{\mathbb{R}^n} \varphi(x) d\mu(x) + \int_{\mathbb{R}^n} \psi(y) d\nu(y)$$

$\phi_c = \{(\varphi, \psi) \in (C^0(\mathbb{R}^n))^2 \text{ tq } \varphi(x) + \psi(y) \leq C(x, y) \text{ presque partout} \}$

Alors,  $\inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = \sup_{(\varphi, \psi) \in \phi_c} J(\varphi, \psi)$

**Note:-**

Interprétation :

1. On embauche un transporteur.
2. Il achète de la masse située en x au prix  $\varphi(x)$ .
3. Il vous débarasse au prix  $\int \varphi(x) d\mu(x)$
4. Il vous revend de la masse en y au prix  $\psi(y)$
5. On rachète  $\nu$  au prix  $\int \psi(y) d\nu(y)$

On embauche le transporteur sous la condition :  $\varphi(x) + \psi(y) \leq C(x, y)$

**Proof:** D'une part :

Soit  $(\varphi, \psi) \in \phi_c$

Soit  $\Pi \in \Pi(\mu, \nu) | I[\Pi] = \tau(\mu, \nu)$

$$J(\varphi, \psi) = \int_{\mathbb{R}^n} \varphi(x) d\mu(x) + \int_{\mathbb{R}^n} \psi(y) d\nu(y) \leq \int C(x, y) d\Pi(x, y) = I[\Pi] = \tau(\mu, \nu) = \inf_{\Pi} I[\Pi]$$

Ainsi,  $J(\varphi, \psi) \leq \inf_{\Pi} I[\Pi]$

D'autre part :

**Note:-**

Remarque :

Cette correspondance permet de résoudre le problème primal en utilisant des techniques de programmation linéaire plutôt que des techniques de programmation non linéaire, ce qui peut rendre les calculs plus efficaces.

**Definition 1.4.1: Les fonction C-concaves (relatif au coût)**

Soit,  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty\}$

On définit sa fonction C-conjuguée par :

$$\varphi^c : \begin{cases} \mathbb{R}^n & \longrightarrow \mathbb{R} \cup \{-\infty\} \\ y & \longmapsto \inf_{x \in \mathbb{R}^n} (C(x, y) - \varphi(x)) \end{cases}$$

On dit que  $\varphi$  est C-concave si  $\exists \psi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty\}$  tq  $\forall x \in \mathbb{R}^n, \varphi(x) = \psi^c(x)$

**Lemme 1.4.1**

1.  $\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \varphi(x) + \varphi^c(y) \leq C(x, y)$
2.  $\varphi^c = \varphi^{ccc}$
3.  $\varphi = \varphi^{cc} \Leftrightarrow \varphi$  est C-concave

**Proof:** 1. Par définition,  $\forall y, \varphi^c(y) = \inf_{x \in \mathbb{R}^n} (C(x, y) - \varphi(x)) \leq C(x, y) - \varphi(x) \forall x \in \mathbb{R}^n$   
Ainsi,  $\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \varphi(x) + \varphi^c(y) \leq C(x, y)$

2. Par définition,  $\varphi^{cc}(y) = \inf_{x \in \mathbb{R}^n} (C(x, y) - \varphi^c(x)) \geq \varphi(y)$   
En effet comme démontré ci-dessus,  $\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \varphi(y) \leq C(x, y) - \varphi^c(x)$

$\Rightarrow$  D'une part,  $\varphi^c(x) = \inf_y (C(x, y) - \varphi(y)) \geq \inf_y (C(x, y) - \varphi^{cc}(y)) = \varphi^{ccc}(x)$   
Ainsi,  $\varphi^c(x) \geq \varphi^{ccc}(x)$

$\Rightarrow$  Et d'autre part,  $\inf_{x \in \mathbb{R}^n} (C(x, y) - \varphi^c(x)) \leq C(x, y) - \varphi^c(x) \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$   
Donc,  $\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \varphi^c(x) \leq C(x, y) - \varphi^{cc}(x)$   
Donc,  $\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \varphi^c(x) \leq \inf_{x \in \mathbb{R}^n} (C(x, y) - \varphi^{cc}(x)) = \varphi^{ccc}(x)$   
Ainsi,  $\varphi^c(x) \leq \varphi^{ccc}(x)$

Ce qui montre bien que,  $\varphi^c(x) = \varphi^{ccc}(x)$

3.  $\Rightarrow$  Si  $\varphi$  est C-concave alors  $\exists \psi | \varphi = \psi^c$   
Donc,  $\varphi^{cc} = \psi^{ccc} = \psi^c = \varphi$   
 $\Leftarrow$  Si  $\varphi = \varphi^{cc}$ , alors  $\varphi = (\varphi^c)^c$  Donc  $\varphi$  est C-concave.

☺

**Théoreme 1.4.2**

La dualité de Kantorovitch peut être restreinte à des couples de fonction C-conjuguées.

$$\sup_{(\varphi, \psi) \in (C(\mathbb{R}^n)^2)} J(\varphi, \psi) = \max_{(\psi^c, \psi)} J(\psi^c, \psi)$$

**Proof:** On montre que le sup est un max.

☺

**Corollaire 1.4.1 Les plans de transferts optimaux sont caractérisés par leur support**

Si  $(\varphi, \psi)$  est un maximiseur pour le problème de Kantorovitch dual, alors

$\pi \in \Pi(\mu, \nu)$  est un minimiseur pour le problème de Kantorovitch primal si et seulement si  $\pi$  est concentrée sur  $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \varphi(x) + \psi(y) = C(x, y)\}$

### 1.4.2 Application(s)

**Proposition 1.4.1** On considère un objet indexé par  $j$  présent en quantité  $\nu_j$ .

$(\text{type}(j), \text{quantité}(j)) = (j, \nu_j)$

On considère un consommateur indexé par  $i$  présent en quantité  $\mu_i$ .

$(\text{type}(i), \text{quantité}(i)) = (i, \mu_i)$

Utilité de l'objet  $j$  pour l'agent  $i$ .

**Hypothèse** L'utilité est transférable.

L'objet  $j$  a une utilité nette  $U_{r,j} - P_j$ .

Pour un système de prix  $P_j$ , l'agent  $i$  choisit l'objet  $j_p$  qui maximise  $U_{ij} - P_j$ .

Transfert optimal:  $SUP_{\Pi} \sum_{ij} U_{ij} \Pi_{ij}$  sous la contrainte  $\sum_j \Pi_{ij} = \mu_i ; \sum_i \Pi_{ij} = \nu_j$

#### Note:-

**Explication :**

$$\mu = \sum_i \mu_i \delta_{x_i}$$

$$\nu = \sum_j \nu_j \delta_{y_j}$$

$$C(i, j) = -u_{i,j}$$

$$I[\Pi] = - \sum_{ij} u_{ij} \Pi_{ij}$$

**Utilité maximale :**  $INF_{\Pi} I[\Pi] = SUP_{\Pi} \{- \sum_i \varphi_i \mu_i - \sum_j \psi_j \nu_j\}$

**Problème dual :**  $(D) : INF_{P_j} \{\sum_j \nu_j P_j + \sum_i \mu_i MAX_j (u_{ij} - P_j)\}$

#### Definition 1.4.2: Prix d'équilibre

Un système de prix qui satisfait (D) est un prix d'équilibre du problème. Un tel système de prix permet d'atteindre l'optimum global

#### Note:-

Pour le mar. 29 nov. 2022

Lire Guillaume Carlier, Teaching, Transfert Optimal (Chapitre 3 Matching equilibre)

Faire exercice 5, 6, 8 et 9.

## 1.5 La distance de Wasserstein

### Note:-

Introduction :

La distance de Wasserstein, également appelée distance de Monge-Kantorovich, est utilisée en théorie du transport optimal pour mesurer la distance entre deux distributions de probabilité. Elle est basée sur le principe de la minimisation du coût total de transport entre les deux distributions. Elle a pour avantage de permettre de prendre en compte les informations sur les coûts de transport entre les différents points d'origine et de destination.

### Definition 1.5.1: Distance de Wasserstein

On munit  $\mathbb{R}^n$  d'une distance  $d$ . On considère la fonction cout :  $C(x, y) = [d(x, y)]^p$  Par convention, si  $p = 0$ , par convention,  $(d(x, y))^0 = \begin{cases} = 0 & \text{si } x = y \\ = 1 & \text{sinon} \end{cases}$

### Definition 1.5.2: Espace de Wasserstein

L'espace de Wasserstein d'ordre  $p$  associé à  $\mathbb{R}^n$  est défini comme suit:

$$\mathcal{P}_p(\mathbb{R}^n) = \{\mu | \mu \text{ est une mesure de probabilité telle que } \int_{\mathbb{R}^n} d(x_0, x)^p d\nu(x) < \infty\}$$

### Note:-

Interpretation:

L'espace de Wasserstein est un espace de probabilités mesurables qui est défini comme l'ensemble des distributions de probabilité qui peuvent être reliées par des plans de transfert optimaux.

### Note:-

Remarque:

Si  $p \in \mathcal{P}_p(\mathbb{R}^n)$  alors  $\forall y \in \mathbb{R}^n, \int d(y, x)^p d\mu(x) < \infty$

Si  $d$  est borné  $\mathcal{P}_p(\mathbb{R}^n) = \mathcal{P}(\mathbb{R}^n)$

### Definition 1.5.3

Soit  $p \geq 1$

On définit,  $W_p^p(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} d(x, y)^p d\pi(x, y)$

### Note:-

On remarque que,  $W_p(\mu, \nu) = [W_p^p(\mu, \nu)]^{\frac{1}{p}}$

### Théoreme 1.5.1

$W^p$  est une mesure

*Proof:* .

#### Definition 1.5.4: Inégalité Triangulaire

Soient  $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_p(\mathbb{R}^n)$   $\pi_{1,2}(\mu_1) = \mu_2$   $\pi_{2,3}(\mu_2) = \mu_3$

Des plans de transports optimaux. On note  $\pi$  la mesure de proba sur  $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$  définie dans le lemme définit ci-dessous.

On note  $\pi_{1,3}$  sa marginale sur  $\mathbb{R}^n \times \mathbb{R}^n$

On a  $\pi_{1,3} \in \Pi(\mu_1, \mu_3)$  mais pas (forcément) optimal.

##### Note:-

$\Pi$  à pour marginale  $\pi_{1,3}$  sur  $\mathbb{R}_1^n \times \mathbb{R}_2^n$

$$\forall \varphi \in (\mathbb{R}^n \times \mathbb{R}^n) \int \varphi(x, y) d\Pi(x, y, z) = \int \varphi(x, y) d\pi_{1,3}(x, y)$$

La marginal d'une mesure est une projection d'une mesure sur un sous-espace de dimension inférieure. Elle est utilisée pour décrire les propriétés d'une mesure en se concentrant sur une sous-partie de ses dimensions.

⊕

#### Exemple 1.5.1

$p = 2$

$W_2$  est la distance de Wasserstein quadratique.

Si  $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ ,  $a \in \mathbb{R}^n$

$$W_2(\mu, \delta_a) = \int_{\mathbb{R}^n} (x - a)^2 d\mu(x)$$

Et la moyenne de  $\mu$  (son espérance) est définie comme :  $m = \int x d\mu(x) dx$

$$m = \text{INF}_{a \in \mathbb{R}^n} W_2(\mu, \delta_a)$$

##### Note:-

Cela permet de définir des espérances sur des espaces compliqués dans lesquels on ne peut pas définir d'intégrale.

#### Exemple 1.5.2

$p = 1$

$\tau_1$  s'appelle la distance de Rubinstein-Kantorovitch.

#### Definition 1.5.5: CDF

$\mu \in \mathcal{P}(\mathbb{R})$  On définit sa CDF par :  $F(x) = \int_{-\infty}^x d\mu$

#### Definition 1.5.6

On appelle  $F^{-1}$  l'inverse généralisé de  $F$  définie sur  $[0,1]$  par :  $F^{-1}(t) = \text{INF}_{x \in \mathbb{R} | F(x) > t}$

#### Proposition 1.5.1

$$W_1(\mu, \nu) = \int_0^1 F^{-1}(t) - G^{-1}(t) dt$$

$$W_1(\mu, \nu) = \int_{\mathbb{R}} |F(x) - G(x)| dx$$

$$W_1(\mu, \nu) = \|F - G\|_{L_1(\mathbb{R})}$$

##### Note:-

On note,  $\mu_1, \dots, \mu_n \in \mathcal{P}_2(\mathbb{R}^n)$

$\lambda_1, \dots, \lambda_n \in \mathbb{R}^+$  tel que  $\sum_{i=1}^n \lambda_i = 1$

**Definition 1.5.7**

$\mu^* = \inf_{\mu \in \mathcal{P}_2(\mathbb{R})} \sum_{i=1}^n \lambda_i W_2^2(\mu, \mu_i)$   
est le barycentre de  $(\mu_1, \lambda_1) \dots (\mu_n, \lambda_n)$

**Note:-**

Voir cours : Peyré / Cuturi : Computational Transfert optimal.

## 1.6 Estimation numérique du transfert optimal

### 1.6.1 L'algorithme de Sinkhorn

#### Note:-

Introduction :

L'algorithme de Sinkhorn est un algorithme utilisé pour résoudre des problèmes de transport optimal avec des contraintes de positivité et de normalisation sur les plans de transport. Il est basé sur une méthode itérative qui utilise des techniques de programmation linéaire pour optimiser les variables du problème primal et du problème dual, et une technique de régularisation pour garantir la convergence de l'algorithme. Il est particulièrement utile pour résoudre des problèmes de transport optimal avec des coûts de transport qui varient considérablement entre les différents points d'origine et de destination.

#### Le contexte

#### Note:-

Habituellement présenté dans le cas discret

$$a = \sum_{i=1}^n a_i \delta_{x_i} \quad b = \sum_{j=1}^m b_j \delta_{y_j}$$

$C_{ij} = C(x_i, y_j)$  C correspond à une fonction cout

$\tau(a, b) = \inf \{ \sum_{i,j} C_{ij} P_{ij} \mid P \in M_{n \times m}(\mathcal{R}) \}$  P est une mesure de probabilité

$P \mathbb{1}_n = a$  et  $tP \mathbb{1}_m = b$  Avec ,  $\mathbb{1}_m = (1, 1, 1, \dots, 1)$  m fois

#### Régulation entropique

#### Formulation

##### Definition 1.6.1: L'entropie de couplage

L'entropie discrète d'une matrice de couplage (une matrice qui envoi une mesure a sur une mesure b)

$$H(P) = \sum_{i,j} P_{ij} (\log P_{ij} - 1)$$

Avec, P une matrice bistochastique (la somme des éléments de chaque colonne vaut 1)

#### Note:-

Remarque le terme -1 ne sert à rien.

On pourrait avoir :  $H(P) = \sum_{i,j} P_{ij} (\log P_{ij})$

car,

$$\sum_{i,j} P_{ij} = 1$$

$$\int d\pi = 1 \text{ avec } \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$$

##### Proposition 1.6.1

$$H : M_{n \times m}(\mathbb{R}) \rightarrow \mathbb{R}^+$$

$$\nabla H(P) = (\log(P_{ij}) + 1)_{i,j}$$

$$\nabla^2 H(P) = \text{diag}\left(\frac{1}{P_{ij}}\right)$$

$\text{diag}(P_{ij})$  Correspond a une matrice avec tous les termes de P sur la diagonale

**Proof:** Montrons que,  $\nabla H(P) = (\log(P_{ij}))_{i,j}$

$$(P, Q) \in (M_{n \times m}(\mathbb{R}))^2$$

$$H(P + \delta Q) = H(P) + \delta \langle \nabla H(P), Q \rangle + \frac{\delta^2}{2} \langle \nabla^2 H(P) Q, Q \rangle + o(\delta^2) \text{ Formule Taylor}$$

$$H(P + \delta Q) = \sum_{i,j} ((P_{ij} + \delta Q_{ij}) \log(P_{ij} + \delta Q_{ij})) \text{ Par définition}$$

En developpant la formule de la définition on montre qu'elle est égale à la formule de Taylor On la développe à l'aide d'un développement limité sur le log. ☺

**Note:-**

Tout cela permet de montrer que l'entropie est **convexe** sur l'ensemble des matrices bistochastiques (qui est un ensemble convexe)

**Exercice**

Parmi les couplage P, quels sont ceux qui minimisent l'entropie ?

**Solution:** Soit  $P^*$  un min de H sous la contrainte  $\sum_{ij} P_{ij} = 1$

Avec,  $F(P^*) = 0$  ou  $F(P) = \sum_{ij} P_{ij} - 1$  Ce qui correspond à un problème d'optimisation sous contrainte.

Alors,  $\exists \lambda \in \mathbb{R}$  tq  $\nabla H(P^*) + \lambda \nabla F(P^*) = 0$  et,  $F(P^*) = 0$

$$\forall ij (\log P_{ij} + 1 + \lambda) = 0$$

$$\forall ij P_{ij} = \frac{1}{nm}$$

Donc l'entropie est minimal lorsque tous les points de a sont coupler avec tous les points de b.

**Exercice**

$$\tau^\varepsilon = \inf_{P \in \Pi(a,b)} \{ \langle P, C \rangle + \varepsilon H(P) \}$$

C'est un problème d'optimisation (comme pour le problème de transfert optimal) mais la fonction à minimiser est convexe. En effet c'est la somme d'une fonction convexe et d'une fonction linéaire. Minimiser une fonction convexe est beaucoup plus simple que de minimiser une fonction linéaire.

On notera ce problème :  $(T.O^\varepsilon)$

**Théoreme 1.6.1**

La solution  $(P^\varepsilon)$  de  $(T.O^\varepsilon)$

- converge vers le couplage d'entropie minimal :  $a \otimes b = a^t b$  quand  $\varepsilon \rightarrow \infty$
- converge vers  $P^0$  quand  $\varepsilon \rightarrow 0$   
Ou  $P^0 = \operatorname{argmax} \{ H(P) \}$

$$P \in \Pi(a,b) \langle P, C \rangle = \tau^\varepsilon(a,b)$$

**Proof:** On note  $(P^\varepsilon)$  l'unique solution de  $(T.O^\varepsilon)$

$\forall \varepsilon P^\varepsilon \in \Pi(a,b)$  qui est borné

On peut extraire une sous suite  $P^\varepsilon \rightarrow P^0$

$\Pi(a,b)$  est fermé dans  $M_{n \times m}(\mathbb{R})$

Donc,  $P^0 \in \Pi(a,b)$

Fixons un  $P^* \in \Pi(a,b)$

Tel que par définition,  $\langle C, P^* \rangle = \sum_{ij} C_{ij} P_{ij} = \tau_c(a,b)$

On a  $\langle C, P^* \rangle \leq \langle C, P^\varepsilon \rangle \forall \varepsilon$  (En effet  $P^*$  est optimal pour T.O)

D'autre part,  $\langle C, P^* \rangle + \varepsilon H(P^*) \geq \langle C, P^\varepsilon \rangle + \varepsilon H(P^\varepsilon)$

☺

**Réécriture du problème de transfert optimal avec régularisation entropique****Definition 1.6.2: Kullback-Leibler divergence**

$$KL(P|Q) = \sum_{ij} P_{ij} \log\left(\frac{P_{ij}}{Q_{ij}}\right)$$

**Definition 1.6.3: Le rayon de Gibbs**

$\varepsilon \geq 0$  fixé

Le rayon de Gibbs associé au cout c est défini par:  $K_{ij}^\varepsilon = \exp\left(-\frac{C_{ij}}{\varepsilon}\right)$



**Proposition 1.6.2**  $P^\varepsilon = \operatorname{argmin}_{P \in \Pi(a,b)} KL(P \mid K^\varepsilon)$

## L'algorithme de Sinkhorn

**Proposition 1.6.3** L'unique solution de  $(T.O)^\varepsilon$  est de la forme :

$$P^\varepsilon = U^\varepsilon K^\varepsilon V^\varepsilon$$

$$\text{Ou } U^\varepsilon \in \mathbb{R}^n$$

$$V^\varepsilon \in \mathbb{R}^n$$

$$K^\varepsilon = \text{noyau de Gibbs}$$

$$P^\varepsilon \text{ minimise } \langle C, P \rangle + \varepsilon H(P)$$

Sous les contraintes :

- $P \mathbb{1}_n = a$  (n contraintes)
- $tP \mathbb{1}_n = b$  (m contraintes)

On peut le résoudre avec les multiplicateurs de Lagrange

### Note:-

On peut écrire que  $P = UKV$  sous forme matricielle

$$P = \operatorname{diag}(U).K.\operatorname{diag}(V)$$

Ainsi les contraintes  $P \in \Pi(a, b)$  se réécrivent

$$\operatorname{diag}(U)K\operatorname{diag}(V)\mathbb{1}_n = 0$$

$$\text{ici } a = U \odot KV$$

$$\operatorname{diag}(V)^t K \operatorname{diag}(U)\mathbb{1}_n = 0$$

$$\text{ici } b = V \odot^t KU$$

### Definition 1.6.4

$$\frac{U}{W} = \left( \frac{u_1}{w_1}, \frac{u_2}{w_2}, \dots, \frac{u_n}{w_n} \right)$$

## L'algorithme de Sinkhorn

**Solution:** .

$v^0 \in \mathbb{R}^n$  (on le choisit)

$$l = 0$$

$$u^{(l+1)} = \frac{a}{Kv^{(l)}} ; v^{(l+1)} = \frac{b}{tKu^{(l+1)}}$$

### Note:-

- Meilleure convergence quand  $\varepsilon$  grand
- Complexité :  $\frac{n^2 \log(n)}{\tau^3}$

## 1.6.2 Formulation dynamique

### Note:-

Cadre continu

Mesures à densités (i.e des fonctions)

$$\mu_0 = \varphi_0 dx$$

$$\mu_1 = \varphi_1 dx$$

$$C(x, y) = \|x - y\|^2$$

## Théorème de Brenier

### Théoreme 1.6.2 Théorème de Brenier

Si  $\mu_0$  et  $\mu_1$  sont à densité par rapport à Lebesgue, alors il existe un unique plan de transport  $\pi$  et ce plan est supporté sur le graphe  $(x, T(x))$  d'une application de Monge  $T$ .

### Exemple 1.6.1 (Application)

En 1D, si  $\varphi$  est convexe,  $\nabla\varphi = \varphi'$  est croissante  
Ce théorème permet de généraliser la notion en multi dimension.

## L'équation de transport (1D)

### Exercice

Résoudre l'équation de transport suivante :

- $\frac{\partial\varphi}{\partial t}(t, x) + \frac{\partial}{\partial x}(v(t, x)\varphi(t, x)) = 0$
- $\varphi(0, x) = \varphi_0(x)$

Inconnue :  $\varphi$

Donnée :  $v$

**Solution:** La solution de  $\frac{\partial\varphi}{\partial t}(t, x) + c \frac{\partial}{\partial x}(\varphi(t, x)) = 0$   
est  $\varphi(t, x) = \varphi_0(x - ct)$

On peut le vérifier en plaçant la solution dans l'équation de transport.

### Théoreme 1.6.3 Formulation dynamique du transport optimal (Théorème de Benamou/Brenier)

$$W_2^2(\rho_0, \rho_1) = \tau_c^2(\rho_0, \rho_1) = \inf \left\{ \int_0^1 \int |v(t, x)|^2 \varphi(t, x) dx dt \right\}$$

### Note:-

Interpretation du barycentre de Wasserstein de 2 mesures au sens du transfert optimal.

En particulier  $\varphi(\frac{1}{2}, x)$  est la moyenne au sens de Wasserstein entre  $\rho_0$  et  $\rho_1$  une nouvelle méthode numérique pour calculer  $W^2$

# Chapter 2

## TD

### 2.1 TD1

#### Exercice

**Exercice 1.** On considère le coût  $c(x, y) = |x - y|$ . Dans chacun des cas, donner les solutions des problèmes de Monge et de Kantorovitch.

1.  $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \quad \nu = \frac{1}{3}\delta_{-1} + \frac{1}{3}\delta_2 + \frac{1}{3}\delta_3,$
2.  $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \quad \nu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1,$
3.  $\mu = \frac{1}{3}\delta_0 + \frac{1}{3}\delta_1 + \frac{1}{3}\delta_2, \quad \nu = \frac{1}{3}\delta_{-1} + \frac{1}{3}\delta_0 + \frac{1}{3}\delta_3.$

**Solution:**

1.  $0 \geq \alpha \geq \frac{1}{3}; 0 \geq \beta \geq \frac{1}{3}$   
 $\frac{1}{6} \geq \alpha + \beta \geq \frac{1}{2}$   
 $\Pi_{\alpha, \beta} = \alpha\delta_{(0, -1)} + \beta\delta_{(0, 2)} + (\alpha + \beta)\delta_{(0, 3)}$

2. Equivalent à l'exemple du cours.

#### Exercice

**Exercice 2.** Soit  $T : \mathbb{R} \rightarrow \mathbb{R}$  définie par  $T(x) = x + 1$ ,  $S : \mathbb{R} \rightarrow \mathbb{R}$  définie par  $S(x) = 2x$  et  $Z : \mathbb{R} \rightarrow \mathbb{R}$  définie par  $Z(x) = 2 - x$ . On définit  $\mu = \mathbb{1}_{[0, 1]}$  et  $\nu = \mathbb{1}_{[1, 2]}$ . A-t-on  $T\#\mu = \nu$  ?  $S\#\mu = \nu$ ?  $Z\#\mu = \nu$ ?

**Solution:** Soit  $\varphi \in C^0(\mathbb{R})$ ,

$$\int_{[0, 1]} (\varphi \circ T)(x) d\mu(x) = \int_{[0, 1]} \varphi(1 + x) dx = \int_{[1, 2]} \varphi(y) dy = \int \varphi(y) d\nu(y)$$

$\Rightarrow T\#\mu = \nu$

$$\int_{[0, 1]} (\varphi \circ S)(x) d\mu(x) = \int_{[0, 1]} \varphi(2x) dx = \frac{1}{2} \int_{[0, 2]} \varphi(y) dy$$

$\Rightarrow S\#\mu = \frac{1}{2}\mathbb{1}_{[0, 2]} \Rightarrow S\#\mu = \frac{1}{2}(\mu + \nu) \Rightarrow S\#\mu \neq \nu$

$$\int_{[0, 1]} (\varphi \circ Z)(x) d\mu(x) = \int_{[0, 1]} \varphi(2 - x) dx = \int_{[1, 2]} \varphi(y) dy = \int \varphi(y) d\nu(y)$$

$\Rightarrow Z\#\mu = \nu$

#### Exercice

**Exercice 3.** (Non-unicité pour un coût convexe - Book shifting). On définit  $\mu = \frac{1}{2}\mathbb{1}_{[0, 2]}$  et  $\nu = \frac{1}{2}\mathbb{1}_{[1, 3]}$  et le coût  $c(x, y) = |x - y|$ . Soit  $T_1(x) = x + 1$  et

$$T_2(x) = \begin{cases} x + 2, & \text{si } x \in [0, 1], \\ x, & \text{si } x \in (1, 2]. \end{cases}$$

Montrer que  $T_1$  et  $T_2$  sont deux applications optimales.

**Solution:** Vérifions que :

$$\begin{aligned} T_1 \# \mu &= \nu \\ \int_{[0,2]} (\varphi \circ T_1)(x) d\mu(x) &= \frac{1}{2} \int_{[0,2]} \varphi(x+1) dx = \frac{1}{2} \int_{[1,3]} \varphi(y) dy = \int \varphi(y) d\nu(y) \\ I(T_1) &= \int_{\mathbb{R}} C(x, T_1(x)) d\mu(x) = \frac{1}{2} \int_{[0,2]} |x - T_1(x)| dx = \frac{1}{2} \int_{[0,2]} 1 dx = 1 \end{aligned}$$

$$\begin{aligned} T_2 \# \mu &= \nu \\ \int_{[0,2]} (\varphi \circ T_2)(x) d\mu(x) &= \frac{1}{2} \left( \int_{[0,1]} \varphi(x+2) dx + \int_{[1,2]} \varphi(x) dx \right) = \frac{1}{2} \left( \int_{[2,3]} \varphi(x) dx + \int_{[1,2]} \varphi(x) dx \right) = \frac{1}{2} \int_{[1,3]} \varphi(x) dx \\ I(T_2) &= \int_{[0,2]} C(x, T_2(x)) d\mu(x) = \frac{1}{2} \int_{[0,1]} |x - T_2(x)| dx + 0 = \frac{1}{2} \int_{[0,1]} 2 dx = 1 \end{aligned}$$

De façon générale,  $T \# \mu = \nu$  alors :

$$I(T) = \frac{1}{2} \int_{[0,2]} |x - T(x)| dx \geq \frac{1}{2} \left| \int_{[0,2]} |x| dx - \int_{[0,2]} |T(x)| dx \right| = \frac{1}{2} |2 - 4| = 1$$

### Exercice

**Exercice 4.** (Non existence d'une application de transport). On prend  $\mu$  la mesure uniforme sur  $[0, 1]$  et  $\nu$  la mesure uniforme sur  $[-1, 1]$ . On considère le coût  $c(x, y) = (x^2 - y^2)^2$ . 1. Pour tout entier  $n$  on définit l'application

$$T_n(x) = \begin{cases} 2x - \frac{k}{2n}, & \text{pour } x \in \left[\frac{k}{2n}, \frac{k+1}{2n}\right] \text{ si } k \text{ est pair,} \\ -2x + \frac{k+1}{2n}, & \text{pour } x \in \left[\frac{k}{2n}, \frac{k+1}{2n}\right] \text{ si } k \text{ est impair.} \end{cases}$$

Montrer que  $T_n \# \mu = \nu$  et montrer que

$$\lim_{n \rightarrow \infty} \int_0^1 c(x, T_n(x)) d\mu(x) = 0.$$

2. En déduire qu'il n'existe pas d'application de transport qui soit optimale. 3. Construire un plan de transport optimal.

### Exercice

**Exercice 5.** (Transport quadratique et translation). On considère le coût  $c(x, y) = (x - y)^2$  sur  $\mathbb{R}^2$ . Pour  $a \in \mathbb{R}$ , on définit la translation  $\tau_a(x) = x - a$ . Soit  $f$  et  $g$  deux fonctions continues. Le but est de montrer que

$$\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) = \mathcal{T}_c(f, g) + (b - a)^2 + 2(b - a)(m_g - m_f),$$

où

$$m_f = \int_{\mathbb{R}} x f(x) dx, \quad m_g = \int_{\mathbb{R}} x g(x) dx.$$

1. Soit  $T$  une application optimale qui envoie  $f$  sur  $g$ . On définit  $S$  par  $S(x) = T(x - a) + b$ . Montrer que  $S \# (f \circ \tau_a) = g \circ \tau_b$ .

2. Montrer que

$$\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) \leq \int_{\mathbb{R}} |S(x) - x|^2 f(\tau_a(x)) dx$$

3. En déduire que

$$\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) \leq \mathcal{T}_c(f, g) + (b - a)^2 + 2(b - a)(m_g - m_f),$$

4. De même montrer que

$$\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) \geq \mathcal{T}_c(f, g) + (b - a)^2 + 2(b - a)(m_g - m_f),$$

et en déduire (1)

5. En déduire que  $\mathcal{T}_c(\mathbb{1}_{[0,1]}, \mathbb{1}_{[1,2]}) = 1$ .

**Solution:**

- Montrons que :  $S\#(f \circ \tau_a) = g \circ \tau_b$   
On sait que  $T$  est une application optimale de  $f$  vers  $g$ . Donc, :  $T\#f = g$   
Donc,  $\forall \psi$  continue :  $\int (\psi \circ T)(x) df(x) = \int \psi(x) dg(x)$   
Et,  $\tau^c(f, g) = \int_{\mathbb{R}} |x - T(x)|^2 df(x)$   
D'autre part,  $\forall \varphi$  continue :  $\int (\varphi \circ S)(x) d(f \circ \tau_a)(x) = \int \varphi(T(x - a) + b) f(x - a) dx$   
On pose,  $u = x - a$   
 $\int \varphi(T(u) + b) f(u) du = \int \varphi \circ \tau_{-b} \circ T(u) f(u) du = \int \varphi \circ \tau_{-b} \circ T(u) g(u) du = \int \varphi(u + b) g(u) du$   
On pose,  $z = u + b$   
 $\int \varphi(z) g(z - b) dz = \int \varphi(z) g \circ \tau_b dz$   
D'où :  $S\#(f \circ \tau_a) = g \circ \tau_b$
- Montrons que :  $\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) \leq \int_{\mathbb{R}} |S(x) - x|^2 f(\tau_a(x)) dx$   
 $\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) = \int_{\mathbb{R}} |T(x) - x|^2 f(\tau_a(x)) dx \leq \int_{\mathbb{R}} |S(x) - x|^2 f(\tau_a(x)) dx$
- Montrons que :  $\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) \leq \mathcal{T}_c(f, g) + (b - a)^2 + 2(b - a)(m_g - m_f)$ ,  
 $\int_{\mathbb{R}} |S(x) - x|^2 f(x - a) dx = \int_{\mathbb{R}} |T(x - a) + b - x|^2 f(x - a) dx = \int_{\mathbb{R}} |T(y) + b - y - a|^2 f(y) dy$   
 $= \int_{\mathbb{R}} |T(y) - y|^2 f(y) dy + \int_{\mathbb{R}} |b - a|^2 f(y) dy + \int_{\mathbb{R}} 2|T(y) - y| \times |b - a| f(y) dy$   
 $= \mathcal{T}_c(f, g) + (b - a)^2 + 2(b - a)(m_g - m_f)$ ,  
D'où,  $\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) \leq \mathcal{T}_c(f, g) + (b - a)^2 + 2(b - a)(m_g - m_f)$ ,
- $\mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) = \int_{\mathbb{R}} |T(x) - x|^2 f(\tau_a(x)) dx$   
On a,  $\mathcal{T}_c(f, g) = \mathcal{T}_c((f \circ \tau_a) \circ \tau_{-a}, (g \circ \tau_b) \circ \tau_{-b}) \leq \mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) + (b - a)^2 + 2(a - b)(m_{g \circ \tau_a} - m_{f \circ \tau_b})$ ,  
Avec,  $m_{f \circ \tau_a} = \int_{\mathbb{R}} x f(x - a) dx = \int_{\mathbb{R}} (y + a) f(y) dy = a + m_f$   
et,  $m_{g \circ \tau_b} = \int_{\mathbb{R}} x g(x - b) dx = \int_{\mathbb{R}} (y + b) g(y) dy = b + m_g$   
Donc,  $\mathcal{T}_c(f, g) \leq \mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) + (b - a)^2 + 2(a - b)(b + m_g - a - m_f)$ ,  
Donc,  $\mathcal{T}_c(f, g) \leq \mathcal{T}_c(f \circ \tau_a, g \circ \tau_b) + (b - a)^2 - 2(b - a)^2 + 2(a - b)(m_g - m_f)$ ,  
D'où  $\mathcal{T}_c(f, g) + (b - a)^2 + 2(b - a)(m_g - m_f) \leq \mathcal{T}_c(f \circ \tau_a, g \circ \tau_b)$   
Ainsi,  $\mathcal{T}_c(f, g) + (b - a)^2 + 2(b - a)(m_g - m_f) = \mathcal{T}_c(f \circ \tau_a, g \circ \tau_b)$
- Pour,  $f = g = \mathbb{1}_{[0,1]}$  et  $\tau_a = \tau_0$  et  $\tau_b = \tau_1$   
 $\tau_c(\mathbb{1}_{[0,1]}, \mathbb{1}_{[1,2]}) = \tau_c(\mathbb{1}_{[0,1]}, \mathbb{1}_{[0,1]}) + (1 - 0)^2 + 2(1 - 0)(m_g - m_f) = 1$

**Exercice**

**Exercice 6.** (Non unicité des potentiels de Kantorovitch). Montrer que si  $(\varphi, \psi)$  est une paire optimale de potentiel de Kantorovitch, alors, pour tout  $a \in \mathbb{R}$ , la paire  $(\varphi + a, \psi - a)$  l'est aussi.

**Exercice**

**Exercice 7.** (Potentiels de Kantorovitch). Donner au moins une paire optimale de potentiel de Kantorovitch pour le book-shifting problem.

**Exercice**

**Exercice 8.** Pour  $x \in \mathbb{R}$  et  $y \in \mathbb{R}$ , on considère le cout  $c(x, y) = \Psi(|x - y|)$  avec

$$\Psi(z) = \begin{cases} 1 - z, & 0 \leq z \leq 1 \\ z - 1, & z \geq 1. \end{cases}$$

On considère les mesures  $\mu = \frac{1}{2}(\delta_{-1} + \delta_2)$  et  $\nu = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$ . Calculer le coût global du transport optimal entre  $\mu$  et  $\nu$ , et donner l'ensemble des plans de transport optimaux.

**Solution:**  $0 \geq \alpha \geq \frac{1}{3}; 0 \geq \beta \geq \frac{1}{3}$

$$\frac{1}{6} \geq \alpha + \beta \geq \frac{1}{2}$$

$$\Pi_{\alpha,\beta} = \alpha\delta_{(-1,0)} + \beta\delta_{(-1,1)} + (\frac{1}{2} - \alpha - \beta)\delta_{(-1,-1)} + (\frac{1}{3} - \alpha)\delta_{(2,0)} + (\frac{1}{3} - \beta)\delta_{(2,1)} + (\alpha + \beta - \frac{1}{6})\delta_{(2,-1)}$$

$$I[\Pi_{\alpha\beta}] = \beta + \frac{1}{2} - \alpha - \beta + \frac{1}{3} - \alpha + 2\alpha + 2\beta - \frac{1}{3} = \frac{1}{2} + 2\beta$$

$$\Pi(\mu, \nu) = \{\Pi_{\alpha\beta=0} | \forall \alpha, \frac{1}{6} \geq \alpha \geq \frac{1}{3}\}$$

### Exercice

**Exercice 9.** Soit  $A \in M_{n,m}(\mathbb{R}), b \in \mathbb{R}^m, c \in \mathbb{R}^n$ . Montrer que

$$\sup_{Ax \leq b} c \cdot x = \inf_{\substack{y \geq 0 \\ {}^t A y = c}} b \cdot y$$

Notation: on dit que  $x \geq 0$  si toutes ses composantes sont positives, et  ${}^t A$  est la transposée de la matrice  $A$ . Pour  $c \in \mathbb{R}^n$  et  $x \in \mathbb{R}^n$ , on note  $x \cdot c$  le produit scalaire entre  $c$  et  $x$ . Indice: S'inspirer de la preuve de la dualité de Kantorovitch.

**Solution:**  $SUP_{Ax \leq b} \langle C, x \rangle = SUP_{x \in \mathbb{R}^n} \langle C, x \rangle + \begin{cases} 0 & \text{si } Ax \leq b \\ -\infty & \text{sinon} \end{cases}$

$$\text{Or, } \begin{cases} 0 & \text{si } Ax \leq b \\ -\infty & \text{sinon} \end{cases} = INF_{y \geq 0} [-\langle Ax - b, y \rangle]$$

$$\text{Si } Ax \leq b$$

Toutes les composantes de  $-(Ax - b)$  sont positives ou nulles.  $\mathcal{L}y \geq 0$  (si toutes les composantes de  $y$  sont positives ou nulles)  $\langle -(Ax - b), y \rangle \geq 0$  Pour  $y = 0, \langle -(Ax - b), y \rangle = 0$

Si  $Ax \leq b$  est faux

$$\exists i | (Ax - b)_i \geq 0$$

$$\text{On prend : } y = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ n \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

### Exercice

**Exercice 10.** On définit pour  $x \in \mathbb{R}$

$$G_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

Calculer  $W_1(G_\sigma, \delta_0)$ .

### Exercice

**Exercice 11** (Interpolation de McCann). Soient  $\mu_0$  et  $\mu_1$  des mesures à densité  $\rho_0$  et  $\rho_1$  par rapport à la mesure de Lebesgue. On considère le coût quadratique  $c(x, y) = \|x - y\|_2^2$ . Soit  $T$  l'application optimale de Monge. 1. Quel théorème garantit l'existence de  $T$ ? 2. On suppose que  $T$  est inversible. Montrer que  $T^{-1}$  envoie  $\mu_1$  sur  $\mu_0$  de façon optimale. 3. On définit pour  $t \in [0, 1]$

$$T_t(x) = (1 - t)x + tT(x),$$

et

$$\mu_t = T_t \# \mu_0.$$

Montrer que  $(T^{-1})_{1-t}$  envoie  $\mu_1$  sur  $\mu_t$ . 4. Montrer que

$$W_2(\mu_0, \mu_t) \leq t W_2(\mu_0, \mu_1).$$

5. Montrer que

$$W_2(\mu_1, \mu_t) \leq (1-t)W_2(\mu_0, \mu_1).$$

6. Calculer  $W_2(\mu_0, \mu_t)$ . 7. On prend

$$\rho_0(x) = \begin{cases} x+1, & x \in [-1, 0] \\ 1-x, & x \in [0, 1] \\ 0, & \text{sinon,} \end{cases} \quad \rho_1(x) = \begin{cases} 1, & x \in [3, 4], \\ 0, & \text{sinon.} \end{cases}$$

Calculer la densité  $\rho_t$  de  $\mu_t$ .