Transport Optimal Computationnel

D'après Gabriel Peyré

12 novembre 2025

Table des matières

Résumé

mailto:gabriel.peyre@ens.fr Notes de cours sur https://arxiv.org/abs/2505.06589 Syllabus : https://docs.google.com/document/u/0/d/1JlDpcS0tkzX8CSgHlUf13ZHQRWu650EtNLrycT39dxk/mobilebasic

Introduction

L'une des motivations principales du cours est de comparer, en apprentissage statistique, des données sous formes de distributions de probabilités, souvent discrètes (nuages de points). On peut penser au transport optimal comme de l'apprentissage non-supervisé : comment associer une distribution de probabilité paramétrique α_θ à une probabilité observée β sur un groupe de points. Dans ce cours, les lettres grecques sont réservées aux distributions de probabilités et les lettres latines aux points. Pour ce faire, on va faire une association de densité $\min_\theta D(\alpha_\theta, \beta)$ qui prend en compte une métrique d. L'idée étant d'utiliser la métrique d pour définir la métrique D, en utilisant la structure de l'espace sous-jacent pour les données. Il faut voir le transport optimal comme un mécanisme d'élévation de l'espace des données vers un espace de probabilité, de sorte que $D = d$ lorsqu'on considère des diracs.

1 Problème de Monge

1.1 Formulation de Monge

Le problème est a été défini pour le but militaire de construire des murs avec des sacs de sables déplacés par des soldats, par Gaspard Monge, dans un papier à l'académie des sciences.

Commençons par un exemple : si on part d'un ensemble de 3 points $x_1, \ldots, x_3$ et qu'on veut atteindre $y_1, \ldots, y_3$, quelle est la manière optimale de donner une bijection dont le coût du déplacement est minimal ? On cherche $T : x_i \to y_{\sigma(i)}$ pour $\sigma \in \mathfrak{S}_3$.

Définition 1.1 Le problème de Monge est le problème d'optimisation suivant :

$$M = \min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^{n} C_{i,\sigma(i)} \tag{Monge}$$

où $C \in \mathbb{R}^{n \times n}$. On dit que M est l'assignation optimale.

En général, on définit $C_{i,j} = c(x_i, y_j)$, et c sera généralement la distance géodésique sur une variété, ou une puissance de la norme p sur $\mathbb{R}^d$. Il n'y a pas besoin de supposer que x et y appartiennent aux mêmes espaces $\mathcal{X}$ et $\mathcal{Y}$.

Il est clair que le problème de Monge est combinatoirement complexe, puisque $|\mathfrak{S}_n| = n!$. Monge, historiquement, a proposé des liens avec l'optique. L'académie des sciences a proposé un prix a celui qui réussirait a proposer une solution (entendre de nos jours, algorithme polynomial pour la réponse). Aujourd'hui, nous avons un algorithme en $\mathcal{O}(n^3)$, qui est optimal dans le pire cas.

1.1.1 Zoologie de Sous-Problèmes

Dans cette section, on va s'intéresser à quelques sous-problèmes spécifiques, dont le résultat est calculable.

Cas 1-D Ici, on suppose que $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, et que $c(x, y) = h(x - y)$ pour h convexe. En général, on aura $c(x, y) = |x - y|^p$ avec $p \geq 1$. Monge avait étudié le cas $p = 1$ qui est de loin le plus difficile.

Dans ce cas, la solution est l'arrangement croissant : l'application T définit ci-dessus est croissante. On trie les points de gauche à droite et on assigne le plus petit x_i au plus petit y_i et ainsi de suite :

Théorème 1.1 Dans le cas 1-D, si M vérifie :

$$x_i < x_j \Rightarrow y_{M(i)} < y_{M(j)}$$

alors M est une solution de (Monge).

Démonstration. Si la propriété n'est pas vérifiée, alors il existe (i, i') tels que $(x_i - x_{i'})(y_{\sigma(i)} - y_{\sigma(i')}) < 0$ et en composant σ par la transposition $\tau_{i,i'}$ on obtient :

$$h(x_i - y_{\tau(\sigma(i))}) + h(x_{i'} - y_{\tau(\sigma(i'))}) < h(x_i - y_{(\sigma(i)}) + h(x_{i'} - y_{\sigma(i')})$$

par convexité de h. ■

Corollaire 1.1 Dans ce cas, on a une complexité en $\mathcal{O}(n \log n)$, en utilisant un algorithme de tri.

R Dans le cas où h est strictement convexe, tous les arrangements optimaux sont croissants, et donc on a l'unicité de la solution M dans le cas où tous les points sont distincts.

Ce cas intervient notamment dans le cas ou par exemple on veut comparer deux groupes de niveaux, ou l'égalisation en niveaux de gris de deux histogrammes de luminance (balance des blancs).

L'algorithme ne se généralise pas en deux dimensions, puisque les trajectoires ne peuvent pas se recouper pour un arrangement optimal (par l'inégalité du parallélogramme) mais que cette propriété n'implique pas l'optimalité de la solution. Pire, il peut exister un nombre exponentiel de solution non-optimales dans ce cas.

Couplage Dans le cas où on peut modéliser le problème par un problème de couplage de graphes, on peut utiliser l'algorithme hongrois pour obtenir une solution en $\mathcal{O}(n^3)$.

1.2 Formulation de Monge Continue

Ici, on va s'intéresser à des mesures boréliennes $\alpha \in \mathcal{M}(\mathcal{X}), \beta \in \mathcal{M}(\mathcal{Y})$, pour $\mathcal{X}, \mathcal{Y}$ des espaces métriques. Pour des questions de facilité, on supposera $\mathcal{X}, \mathcal{Y}$ compacts, et pour l'optimisation on les supposera de plus complets, généralement des parties de $\mathbb{R}^d$. Lorsqu'on aura besoin de retirer l'hypothèse de compacité, on ne demandera de $\mathcal{X}$ et $\mathcal{Y}$ que d'être des espaces polonais. On suppose de plus que α et β sont des mesures de probabilité (positives et de somme 1). On notera l'espace de mesures de probabilité $\mathcal{M}_+^1(\mathcal{X}) = \mathcal{P}(\mathcal{X})$.

Théorème 1.2 — Riesz-Markov Si $\mathcal{X}$ est un espace polonais, et $\mathcal{C}_c(\mathcal{X})$ est l'ensemble des fonctions continues à support compact sur $\mathcal{X}$ muni de la norme $\|\cdot\|_\infty$:

$$\mathcal{M}(\mathcal{X}) = (\mathcal{C}_c(\mathcal{X}))^\circ$$

Démonstration. On utilise pour identification le produit scalaire : $\langle f, \alpha \rangle = \int f \, \mathrm{d}\alpha$. ■

C'est une généralisation du Théorème de Riesz sur un espace de Hilbert. Quelques exemples :

Discrète $\alpha = \sum a_i \delta_{x_i}$ où $a \in \Delta_n$. On appellera souvent le vecteur a l'histogramme, et le vecteur (δ) le nuage de points. Par linéarité, on a :

$$\int f \, \mathrm{d} \left[\sum a_i \delta_{x_i} \right] = \sum a_i f(x_i)$$

Continue/À densité Si $\alpha = \rho_\alpha \beta$ a densité ρ_α par rapport à β :

$$\int f(x)\,\mathrm{d}\alpha(x) = \int f(x)\rho_\alpha(x)\,\mathrm{d}\beta(x)$$

C'est-à-dire : $\langle \cdot, \alpha\rangle = \langle \cdot \times \rho_\alpha, \beta\rangle$. On notera ceci $\alpha << \beta$.

Le produit scalaire défini ci-dessus induit une norme duale dite norme de variation totale :

$$\|\alpha\|_\star = \|\alpha\|_{TV} = \sup_{\|f\|_\infty \leq 1} \langle f, \alpha\rangle$$

Proposition 1.1 Pour $\alpha, \beta \in \mathcal{M}(\mathcal{X})$:

$$\|\alpha - \beta\|_{TV} = |\alpha - \beta|(\mathcal{X}) = \int_{\mathcal{X}} d(|\alpha - \beta|)(x)$$

Quelques exemples :

Discrète Si $\alpha = \sum a_i \delta_{x_i}$ et $\beta = \sum b_i \delta_{x_i}$ où $a, b \in \Delta_n$ (ici, on suppose que les nuages de points sont égaux, mais pas que les a_i, b_i sont non nuls) :

$$\|\alpha - \beta\|_{TV} = \sum_i |x_i - y_i|$$

Continue/Denses Si $\alpha = \rho_\alpha \mathcal{L}$ et $\beta = \rho_\beta \mathcal{L}$:

$$\|\alpha - \beta\|_{TV} = \int |\rho_\alpha - \rho_\beta|\,\mathrm{d}\mathcal{L}$$

Mesures Singulières Si $\alpha = 0$ quand $\beta \neq 0$ et réciproquement, $\|\alpha - \beta\|_{TV} = 2$

Définition 1.2 — Push-Forward. Si on a une application de transport $T : \mathcal{X} \to \mathcal{Y}$, on peut définir une application $T_\sharp : \mathbb{P}(\mathcal{X}) \to \mathbb{P}(\mathcal{Y})$ définie par :

1. $T_\sharp \delta_x = \delta_{T(x)}$;
2. $T_\sharp$ est linéaire.

De manière équivalente : $\beta = T_\sharp \alpha$ qui vérifie $\beta(B) = \alpha(T^{-1}(B))$.

Proposition 1.2 Si $\beta = T_\sharp \alpha$, on a :

$$\int g\,\mathrm{d}\beta = \int g \circ T\,\mathrm{d}\alpha, \text{i.e. } \mathbb{E}(g(Y)) = \mathbb{E}(g(T(X)))$$

pour $Y = T(X)\tilde{\beta}$ un vecteur aléatoire.

On peut alors réécrire le problème de Monge dans le cas continu :

Définition 1.3 Si $\alpha \in \mathbb{P}(\mathcal{X}), \beta \in \mathbb{P}(\mathcal{Y})$, et c est une fonction de coût :

$$M = \inf_{T:\mathcal{X}\to\mathcal{Y}} \left\{ \int c(x, T(x))\,\mathrm{d}\alpha(x) \,\middle|\, T_\sharp \alpha = \beta \right\} \qquad \text{(Monge)}$$

Proposition 1.3 Dans le cas discret : $T_\sharp \alpha = \beta \Leftrightarrow \exists \sigma \in \mathfrak{S}_n, T(x_i) = y_{\sigma(i)}$.

Théorème 1.3 Si α a une densité par rapport à β, alors il existe T telle que $T_\sharp \alpha = \beta$.

R Dans le cas discret, cela revient à dire que $|supp(\alpha)| \geq |supp(\beta)|$. Attention, cette application T n'est pas unique !

Théorème 1.4 — Brenier, 1991 Si $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ et $c(x,y) = \|x-y\|^2$ (généralisé à une distance géodésique mis à la puissance $1 < p < +\infty$) et si α a une densité par rapport à la mesure de Lebesgue (ou au moins ne donne pas de mesure aux espaces de dimension $< d$) alors il y a un unique transport optimal T solution de (Monge) et T est caractérisé par $T = \nabla\varphi$ où φ est convexe (et ∇ est le gradient riemannien). Il y a un unique $\nabla\varphi$ où φ est convexe et $(\nabla\varphi)_\sharp\alpha = \beta$, c'est le transport optimal.

Corollaire 1.2 Les gradients des fonctions convexes $\mathcal{G}$ sont des transports optimaux.

Proposition 1.4 Pour $T \in \mathcal{G}$, $\langle Tx - Ty, x-y\rangle \geq 0$, les gradients de fonctions convexes sont dans l'ensemble $\mathcal{H}$ des fonctions croissantes.

Dans le cas $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, on a même $\mathcal{G} = \mathcal{H}$. En plus grande dimension, $T(x) = Rx$ où R est une petite rotation n'est pas un gradient mais est croissante. Il est possible que φ ne soit pas différentiable, mais on peut montrer que φ va être différentiable presque partout, et donc nos équations sont à entendre comme des égalités à ensemble de mesure nulle près.

Proposition 1.5 $T(x) = Ax$ est un gradient si et seulement si A est symétrique. $T \in \mathcal{G}$ si et seulement si A est définie positive.

Définition 1.4 La 2-distance de Wasserstein est définie par :

$$W_2^2(\alpha,\beta) = \inf_{T_\sharp\alpha=\beta} \int \|x - Tx\|^2 \, \mathrm{d}\alpha(x)$$

Quelques exemples avec les gaussiennes :

Cas 1D Ici, on suppose $\alpha = \mathcal{N}(m_\alpha, \sigma_\alpha^2)$, $\beta = \mathcal{N}(m_\beta, \sigma_\beta^2)$. En prenant $T(x) = \frac{\sigma_\beta}{\sigma_\alpha}(x - m_\alpha) + m_\beta$, on vérifie bien que $T_\sharp\alpha = \beta$ et en primitivant T, on vérifie bien que c'est le gradient (la dérivée) d'une fonction convexe et c'est donc bien un transport optimal. Le coût du transport est :

$$W_2^2(\alpha,\beta) = \|m_\alpha - m_\beta\|^2 + \|\sigma_\alpha - \sigma_\beta\|$$

Cas d-Dimensionnel Ici, $\alpha = \mathcal{N}(m_\alpha, \Sigma_\alpha)$ et $\beta = \mathcal{N}(m_\beta, \Sigma_\beta)$ où $\Sigma_\alpha = \mathbb{E}((X - m_\alpha)^\mathsf{t}(X - m_\alpha))$ est une matrice définie positive. On suppose que $\Sigma_\alpha > 0$. On prend alors $T(x) = A(x - m_\alpha) + m_\beta$ où $A{\Sigma_\alpha}^\mathsf{t}A = \Sigma_\beta$ pour trouver le transport optimal. Cette équation (dite de Ricatti) a une solution symétrique définie positive. On a de plus :

$$W_2^2(\alpha,\beta) = \|m_\alpha - m_\beta\|^2 + \mathcal{B}^2(\Sigma_\alpha, \Sigma_\beta)$$

où $\mathcal{B}$ est la distance de Bures définie par :

$$\mathcal{B}(A,B) = \mathrm{Tr}(A + B - 2(A^{1/2}BA^{1/2})^{1/2})$$

2 Formulation de Kantorovitch

2.1 Définition Discrète

La formulation de Kantorovitch est une relaxation convexe de la formulation de Monge. Il a obtenu un prix Nobel d'économie pour ceci. Ici, on se limite au cas discret $\alpha = \sum_n \alpha_i \delta_{x_i}$ et $\beta = \sum_m b_j \delta_{y_j}$.

Définition 2.1 Un *couplage* ou un *plan* est une matrice $M \in \mathbb{R}_+^{n\times m}$ qui représente le coût de transport de x_i à y_j et telle que :

$$\sum_j P_{i,j} = a_i \wedge \sum_i P_{i,j} = b_j$$

R C'est la même notion que celle de couplage en probabilité : un vecteur aléatoire sur l'espace produit.

On remarquera que les équations définissant un plan peuvent se mettre sous la forme :

$$P\mathbb{1}_m = a \text{ et } {}^{\mathrm{t}}P\mathbb{1}_n = b$$

On autorise ainsi la division de masse, les problèmes de Kantorovitch devenant des problèmes sur des graphes bipartis.

Définition 2.2 Le polytope des couplages entre α et β est l'ensemble des plans entre α et β :

$$\text{Couplages}(\alpha,\beta) = \left\{P \in \mathbb{R}_+^{n\times m} \,\middle|\, P_{i,j} \geq 0, P\mathbb{1}_m = a \text{ et } {}^{\mathrm{t}}P\mathbb{1}_n = b\right\}$$

Kantorovitch avait fait l'hypothèse très forte que l'économie est linéaire.

Définition 2.3 Le *problème de Kantorovitch* est le problème d'optimisation suivant :

$$P = \text{argmin}_P \left\{\langle C, P\rangle \,|\, P \in \text{Couplages}(\alpha,\beta)\right\} \qquad \text{(Kantorovitch)}$$

où $C \in \mathbb{R}^{n\times m}$ est une matrice de coût. On dit que P est le plan optimal.

C'est un problème de programmation linéaire. En général la méthode du simplexe n'est pas polynomial, mais il existe un type de simplexes pour lequel elle l'est, et est en $\mathcal{O}\left(\left(n^3m + m^3n\right)\log\left(mn\right)\right)$

Proposition 2.1 Il existe toujours une solution, et il existe toujours une solution dite *éparse*, telle que :

$$|\{(i,j) \,|\, P_{i,j} \neq 0\}| \leq n + m - 1$$

Démonstration. La preuve d'existence vient du fait que l'ensemble des couplages est un compact non vide (car $P = a{}^{\mathrm{t}}b$ est un couplage dit *indépendant*). ■

Le cas générique est de plus unique, c'est-à-dire que si C, α, β n'a pas une unique solution, en ajoutant du bruit on retrouve une solution optimale.

2.2 Équivalence à Monge

On s'intéresse ensuite aux matrices de permutation P_n, dans le cas $n = m$. On cherche à résoudre le problème non-convexe $\min_{P\in P_n} \langle C, P\rangle$. Clairement, si $\mathcal{B}_n$ est l'ensemble convexe des matrices bistochastiques (l'ensemble des couplages !) :

$$\min_{P\in\mathcal{B}_n} \langle C, P\rangle \leq \min_{P\in P_n} \langle C, P\rangle$$

Définition 2.4 L'ensemble des points extrême d'un convexe C est :

$$\text{Extr}(C) = \left\{P \,\middle|\, \forall (Q,R) \in C^2, P = \frac{Q+R}{2} \Rightarrow Q = R\right\}$$

Théorème 2.1 Si C est compact, $\mathrm{Extr}(C) \neq \varnothing$.

Théorème 2.2 — Krain-Millman Si C est un compact convexe, alors $C = \mathrm{Hull}\left(\mathrm{Extr}\left(C\right)\right)$.

Proposition 2.2 Si C est compact :

$$\mathrm{Extr}(C) \cap \left(\mathrm{argmin}_{p \in C} \left\langle C, P \right\rangle\right) \neq \varnothing$$

Démonstration. En notant que l'ensemble de droite est convexe et compact, on trouve $\mathrm{Extr}(S) \neq \varnothing$ et que $\mathrm{Extr}(S) \subseteq \mathrm{Extr}(C)$. ■

Théorème 2.3 — Birkhoff - von Neumann On a $\mathrm{Extr}(\mathcal{B}_n) = P_n$

Démonstration. On montre d'abord $P_n \subset \mathrm{Extr}(\mathcal{B}_n)$. Cela découle de $\mathrm{Extr}([0,1]) = \{0,1\}$. Si $P \in P_n$, est telle que $P = (Q+R)/2$ avec $Q_{i,j}, R_{i,j} \in [0,1]$, puisque $P_{i,j} \in \{0,1\}$ alors nécessairement $Q_{i,j} = R_{i,j} \in \{0,1\}$.

Montrons maintenant $\mathrm{Extr}(\mathcal{B}_n) \subset P_n$ en montrant que $P_n^c \subset \mathrm{Extr}(\mathcal{B}_n)^c$ avec le complémentaire considéré dans $\mathcal{B}_n$. Choisir $P \in \mathcal{B}_n \setminus P_n$ revient à choisir $\mathbb{P} = (Q+R)/2$ où Q, R sont des matrices bistochastiques distinctes. P définit un graphe biparti de taille $2n$. Le graphe est composé d'arêtes isolées quand $P_{i,j} = 1$ et d'arêtes connectées quand $0 < P_{i,j} < 1$. Si i est un tel sommet à gauche (j à droite), puisque $\sum_j P_{i,j} = 1$, il y a deux arêtes (i, j_1) et (i, j_2) en sortant (de même, (i_1, j) et (i_2, j) entrant en j). On peut donc toujours extraire un cycle par récurrence de la forme :

$$(i_1, j_1, i_2, j_2, \ldots, i_p, j_p)\,, \quad \text{i.e.} \quad i_{p+1} = i_1.$$

On suppose que ce cycle est le plus court de l'ensemble fini de cycle. On a toujours :

$$0 < P_{i_s, j_s}, P_{i_{s+1}, j_s} < 1.$$

Les $(i_s)_s$ et $(j_s)_s$ sont distincts puisque le cycle est le plus court. On pose :

$$\varepsilon = \min_{0 \leq s \leq p} \left\{P_{i_s,j_s}, P_{j_s,i_{s+1}}, 1 - P_{i_s,j_s}, 1 - P_{j_s,i_{s+1}}\right\}$$

c'est-à-dire $0 < \varepsilon < 1$. En séparant le graphe en deux ensembles d'arêtes :

$$\mathcal{A} = \{(i_s, j_s)\}_{s=1}^{p} \quad \text{et} \quad \mathcal{B} = \{(j_s, i_{s+1})\}_{s=1}^{p}\,.$$

On pose Q et R telle que :

$$Q_{i,j} = \begin{cases} P_{i,j} & \text{si } (i,j) \notin \mathcal{A} \cup \mathcal{B}, \\ P_{i,j} + \varepsilon/2 & \text{si } (i,j) \in \mathcal{A}, \\ P_{i,j} - \varepsilon/2 & \text{si } (i,j) \in \mathcal{B}, \end{cases} \quad \text{et} \quad R_{i,j} = \begin{cases} P_{i,j} & \text{si } (i,j) \notin \mathcal{A} \cup \mathcal{B}, \\ P_{i,j} - \varepsilon/2 & \text{si } (i,j) \in \mathcal{A}, \\ P_{i,j} + \varepsilon/2 & \text{si } (i,j) \in \mathcal{B}, \end{cases}.$$

Par définition d'ε, on a $0 \leq Q_{i,j}, R_{i,j} \leq 1$. Puisque chaque arête gauche de $\mathcal{A}$ a une arête droite dans $\mathcal{B}$, (et réciproquement) la contrainte de somme sur les lignes (et sur les colonnes) est maintenue, donc $Q, R \in \mathcal{B}_n$. Finalement, on trouve : $P = (Q+R)/2$. ■

Corollaire 2.1 Pour $m = n$ et $a = b = \mathbb{1}_n$, il existe une solution optimale pour le problème Kantorovitch, qui est une matrice de permutation associée à une permutation optimale pour le problème Monge.

3 Métrique de Wasserstein

3.1 Formulation continue de Kantorovitch

On se place ici dans le cadre continu, où α et β sont des mesures de probabilité arbitraires sur $\mathcal{X}$, $\mathcal{Y}$. On notera P_1 et P_2 les projections

Définition 3.1 Un *couplage* entre α et β est une mesure de probabilité $\pi \in \mathcal{M}_+^1(\mathcal{X}\times\mathcal{Y})$ telle que $P_1\sharp\pi = \alpha$ et $P_2\sharp\pi = \beta$. On note $\mathcal{U}(\alpha,\beta)$ l'ensemble des couplages entre α et β.

En prenant α et β discrètes, on vérifie bien qu'on retrouve la forumlation de l'Equation Kantorovitch.

R Dans le cas où $\mathcal{U}(\alpha,\beta)$ est non vide, le produit tensoriel $\alpha\otimes\beta$ est un couplage, dit *indépendant*.

Définition 3.2 La formulation de Kantorovitch est le problème d'optimisation suivant :

$$\mathcal{L}_c(\alpha,\beta) = \inf_{\pi\in\mathcal{U}(\alpha,\beta)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\mathrm{d}\pi(x,y) = \inf_{(X,y)\sim(\alpha,\beta)} \mathbb{E}[c(X,Y)] \qquad \text{(Kantorovitch)}$$

où $c : (\mathcal{X}\times\mathcal{Y}) \to \mathbb{R}^+$ est la fonction de coût.

Proposition 3.1 On se place dans le cas $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $c(x,y) = \|x-y\|^2$. Si α a densité par rapport à la mesure de Lebesgue, avec $T = \nabla\varphi$ l'application optimale pour Monge, $\pi = (\mathrm{Id}, T)\sharp\alpha$ est l'application optimale pour le problème de Kantorovitch.

3.2 Propriété Métriques du Transport

Ici, on se place dans le cas $\mathcal{X} = \mathcal{Y}$ et on se donne une métrique d sur $\mathcal{X}$. On supposera $c = d^p$ pour un certain $p \geq 1$.

Définition 3.3 Pour $\alpha,\beta \in \mathcal{P}_p(\mathcal{X})$, on définit la distance p-Wasserstein comme :

$$\mathrm{W}_p(\alpha,\beta) = (\mathcal{L}_{d^p}(\alpha,\beta))^{1/p} = \left(\inf_{\pi\in\mathcal{U}(\alpha,\beta)} \int d^p(x,y)\,\mathrm{d}\pi(x,y)\right)^{\frac{1}{p}}$$

R On a une version définie à partir du problème de Monge :

$$\inf_{T\sharp\alpha=\beta} \int d^2(x,T(x))\,\mathrm{d}\alpha(x)$$

mais c'est seulement une pseudo-distance.

Démonstration. Il reste à prouver que W_p définit bien une distance. Pour ça, on a simplement besoin de la séparation et de l'inégalité triangulaire.

Séparation On a : $\mathrm{W}_p(\alpha,\beta) = 0 \Rightarrow \int d(x,y)\,\mathrm{d}\pi^*(x,y) = 0 \Rightarrow d(x,y) = 0, \pi^*$ presque surement. Ceci implique que π^* est supporté sur la diagonale de $\mathcal{X}^2$, c'est-à-dire $P_1\sharp\pi^* = \lambda = P_2\sharp\pi^*$ et donc $\alpha = \beta$.

Inégalité Triangulaire On ne donne la preuve que dans le cas discret, $\alpha = \sum_i a_i\delta_{x_i}$, $\beta = \sum_j b_j\delta_{y_j}$, $\gamma = \sum_k c_k\delta_{z_k}$. En prenant $\pi_{\alpha\beta}$ et $\pi_{\beta\gamma}$ des plans optimaux, on pose :

$$S_{ijk} = \frac{\pi_{\alpha\beta}(i,j)\pi_{\beta\gamma}(j,k)}{b_j}$$

si $b_j \neq 0$, 0 sinon. On a alors $\sum_i S_{ijk} = \pi_{\beta\gamma}(j,k)$ et $\sum_k S_{ijk} = \pi_{\alpha\beta}(i,j)$. Les trois marginales de S sont (α, β, γ). On définit alors :

$$\pi_{\alpha\gamma} = \sum_{i,k} \underbrace{(\sum_j S_{ijk})}_{\pi_{\alpha\gamma}(i,k)} \delta_{x_i, z_k}$$

On vérifie aisément que c'est bien un couplage entre α et γ. On a alors :

$$\begin{aligned}
W_p(\alpha,\gamma) &\leq \left(\sum_{i,k} \pi_{\alpha\gamma}(i,k) d\left(x_i, z_k\right)^p \right)^{1/p} \\
&= \left(\sum_{i,j,k} S_{ijk} d\left(x_i, z_k\right)^p \right)^{1/p} \\
&\leq \left(\sum_{i,j,k} S_{ijk} \left(d\left(x_i, y_j\right) + d\left(y_j, z_k\right)\right)^p \right)^{1/p} \\
&\leq \left(\sum_{i,j,k} S_{ijk} d\left(x_i, y_j\right)^p \right)^{1/p} + \left(\sum_{i,j,k} S_{ijk} d\left(y_j, z_k\right)^p \right)^{1/p} \\
&= \left(\sum_{ij} \pi_{\alpha\beta}(i,j) d\left(x_i, y_j\right)^p \right)^{1/p} + \left(\sum_{jk} \pi_{\beta\gamma}(j,k) d\left(y_j, z_k\right)^p \right)^{1/p} \\
&= W_p(\alpha,\beta) + W_p(\beta,\gamma)
\end{aligned}$$

Pour étendre la preuve à des mesures générales, on utilise le lemme de recollage.

■

Cette structure permet de munir l'espace $\mathcal{P}_p(\mathcal{X})$ d'une structure d'espace métrique. Dans la suite, on va supposer que $\mathcal{X}$ est compact, mais les définitions peuvent être étendues à $\mathbb{R}^d$ par exemple.

Proposition 3.2 Si $\mathcal{X}$ est borné, pour $1 \leq p \leq q$, on a :

$$\mathrm{W}_p(\alpha,\beta) \leq \mathrm{W}_q(\alpha,\beta) \leq (\mathrm{diam}\mathcal{X})^{\frac{q-p}{q}} \mathrm{W}_p(\alpha,\beta)^{p/q}$$

Démonstration. En posant $\varphi(s) = s^{q/p}$ (qui est convexe), par l'inégalité de Jensen :

$$\mathrm{W}_p(\alpha,\beta)^q \leq (\int d(x,y)^p \,\mathrm{d}\pi(x,y))^{q/p} \leq \int d(x,y)^q \,\mathrm{d}\pi(x,y)$$

Donc :

$$\mathrm{W}_p(\alpha,\beta) \leq \inf(\int d(x,y)^q \,\mathrm{d}\pi(x,y))^{1/q} = \mathrm{W}_q$$

La preuve est la même pour l'autre inégalité, en utilisant $d^q \leq (\mathrm{diam}\mathcal{X})^{q-p} d^p$. ■

R La propriété précédente montre que sur un compact, toutes les distances p-Wasserstein définissent la même topologie.

Définition 3.4 Une suite (α_n) converge $\star$-faiblement vers α, noté $\alpha_n \xrightarrow{\star} \alpha$ dans $\mathcal{M}_1^+(\mathcal{X})$ si pour tout f :

$$\int f \,\mathrm{d}\alpha_n \to \int f \,\mathrm{d}\alpha$$

R

- Dans le cas où $\alpha_n = \delta_{x_n}$, on a convergence vers $\alpha = \delta_x$ si et seuleement $x_n \to x$.
- La convergence $\star$-faible est la convergence en loi pour les variables aléatoires.

Définition 3.5 La topologie forte sur $\mathcal{M}_+^1(\mathcal{X})$ est celle définie par la distance de variation totale.

Proposition 3.3 Si on prend d la distance 0-1, alors $\mathrm{W}_p^p(\alpha,\beta) = \frac{1}{2}\|\alpha-\beta\|_{TV}$.

Proposition 3.4 Si $\mathcal{X}$ est compact, alors α_n converge $\star$-faiblement si et seulement si $W_p(\alpha_n,\alpha) \xrightarrow[n\to+\infty]{} 0$. W_p métrise donc la convergence $\star$-faible. Si $\mathcal{X}$ n'est pas compact, la convergence pour W_p équivaut à la convergence $\star$-faible et à la convergence des p-ème moments.

La convergence des sommes de Riemann est équivalente à la convergence $\star$-faible de $\frac{1}{n}\sum_k \delta k/n \to \mathcal{U}_{[0,1]}$.

3.3 Interpolation de McCann

On considère deux mesures de probabilité α, β sur $\mathcal{X} = \mathbb{R}^d$.

Définition 3.6 Pour π un plan optimal entre α et β, et pour $t \in [0,1]$, on définit :

$$\pi_t = P_t \sharp \pi \text{ où } P_t(x,y) = (1-t)x + ty$$

C'est une homotopie, qu'on appelle *Interpolation de McKann* ou *Interpolation de Déplacement*.

Proposition 3.5 L'interpolation de McCann vérifie :

$$\alpha_t \in \mathrm{argmin}_\rho (1-t)\,\mathrm{W}_2^2(\alpha,\rho) + t\,\mathrm{W}_2^2(\beta,\rho)$$

Proposition 3.6 L'espace $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$ est un espace géodésique.

Démonstration. Si T est l'application optimale de Monge pour α, β :

$$\begin{aligned}
\mathrm{W}_2^2(\alpha_t,\alpha_s) &= \int \|(1-t)x + tT(x) - (1-s)x - sT(x)\|^2 \,\mathrm{d}\alpha(x) \\
&= \int \|(t-s)(T(x)-x)\|^2 \,\mathrm{d}\alpha(x) \\
&= |t-s|^2 \, W_2(\alpha,\beta)
\end{aligned}$$

■

4 Gaussiennes et Transport Optimal, Dualité

4.1 Dualité pour les mesures discrètes

Proposition 4.1 Si $\alpha = \sum a_i \delta_{x_i}$ et $\beta = \sum b_j \delta_{y_j}$, et C est une matrice $n \times m$ de coût, en notant $U(\alpha, \beta)$ l'ensemble des couplages entre α et β :

$$L_c(\alpha, \beta) = \max_{(u,v) \in R(\alpha,\beta)} \langle u, \alpha \rangle + \langle v, \beta \rangle$$

où $R(\alpha, \beta) = \{(u, v) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, u \oplus v \leq C\}$.

Démonstration. On a $L_c(\alpha, \beta) = \min_{\pi \in U(\alpha,\beta)} \langle c, \pi \rangle$. On introduit donc :

$$\max_{(u,v) \in \mathbb{R}^n \times \mathbb{R}^m} \langle u, a - \pi \mathbb{1}_m \rangle + \left\langle v + b - {}^{\mathrm{t}}\pi \mathbb{1}_n \right\rangle = \begin{cases} 0 & \text{si } \pi \in U(\alpha, \beta) \\ +\infty & \text{sinon} \end{cases}$$

On a donc :

$$L_c(\alpha, \beta) = \min_{\pi \geq 0} \max_{(u,v) \in \mathbb{R}^n \times \mathbb{R}^m} L(\pi, u, v)$$

avec :

$$\underset{\text{Lagrangien}}{L(\pi, u, v))} \langle c, \pi \rangle + \langle u, \alpha - \pi \mathbb{1}_m \rangle + \left\langle v + \beta - {}^{\mathrm{t}}\pi \mathbb{1}_n \right\rangle$$

Puisque dans le cadre de la programmation linéaire, avoir des solutions suffit pour avoir la dualité forte :

$$\begin{aligned} L_c(\alpha, \beta) &= \max_{u,v} \min_{\pi \geq 0} L(\pi, u, v) \\ &= \max_{u,v} \left(\langle u, \alpha \rangle + \langle v, \beta \rangle + \min_{\pi \geq 0} \langle c, \pi \rangle - \langle u, \pi \mathbb{1}_m \rangle - \left\langle v, {}^{\mathrm{t}}\pi \mathbb{1}_n \right\rangle \right) \\ &= \begin{cases} 0 & \text{si } u \oplus v \leq C \\ -\infty & \text{sinon} \end{cases} \end{aligned}$$

Finalement :

$$\boxed{L_c(\alpha, \beta) = \max_{(u,v) \in R(\alpha,\beta)} \langle u, \alpha \rangle + \langle v, \beta \rangle}$$

ce qui conclut la preuve. ■

R

- Si π est optimal, son support est inclus dans $\{(i, j) \,|\, u_i + v_j = C_{i,j}\}$
- $L_c(\alpha, \beta)$ est convexe en α, β (dual) mais est concave en C.

4.2 Dualité pour des mesures quelconques

Proposition 4.2 Si $\mathcal{X}, \mathcal{Y}$ sont compacts, alors :

$$L_C(\alpha, \beta) = \sup_{(f,g) \in R(C)} \int_{\mathcal{X}} f \, \mathrm{d}\alpha + \int_{\mathcal{Y}} g \, \mathrm{d}\beta$$

où :

$$R(C) = \{(f, g) \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{Y}) \,|\, \forall (x, y), f(x) + g(y) \leq C(x, y)\}$$

On dit que f, g sont des *potentiels de Kantorovitch*.

Dans le cas où $\mathcal{X}$ n'est pas compact, on remplace $\mathcal{C}$ par $\mathcal{C}_b$ et le résultat tient.

R

- Si π est optimal, de même, son support est tel que les potentiels sont égaux au coût.
- Dans le cas où α et β on retrouve le résultat discret.

4.3 c-transformations

On cherche à résoudre :

$$\max_{f\oplus g\leq c} \int g \,\mathrm{d}\beta, \quad f \text{ fixée}$$

On veut donc prendre g aussi grande que possible en ayant :

$$\begin{aligned} g(y) &\leq c(x,y) - f(x) \\ &\leq \inf_{x\in\mathcal{X}} c(x,y) - f(x) \end{aligned}$$

Définition 4.1 Étant donnée $f : \mathcal{X} \to \bar{\mathbb{R}}$, sa *$c$-transformée* est :

$$\begin{array}{rccc} f^c : & \mathcal{Y} & \to & \bar{\mathbb{R}} \\ & f^c(y) & \mapsto & \inf_{x\in\mathcal{X}} c(x,y) - f(x) \end{array}$$

La *$\bar{c}$-transformée* de $g : \mathcal{Y} \to \bar{R}$ est :

$$g^{\bar{c}}(x) = \inf_{y\in\mathcal{Y}} c(x,y) - g(y)$$

Si c est symmétrique, alors $f^c = f^{\bar{c}}$.

R La transformation $(f,g) \to (f, f^c)$ remplace les potentiels duaux par de meilleurs. De même pour $(f,g) \to (g^{\bar{c}}, g)$ et $(f,g) \to (g^{\bar{c}}, f^c)$.

Proposition 4.3 Si c est L-lipschitzienne en y, alors f^c est lipschitzienne.

Démonstration. Exercice. ■

À ce stade, on aurait envie d'itérer à partir de potentiels de base puis d'utiliser les c-transformées pour obtenir un meilleur résultat. La proposition ci-dessous montre que malheureusement ceci ne fonctionnera pas.

Proposition 4.4 Si on note $f^{c\bar{c}} = (f^c)^{\bar{c}}$ alors :

1. $f \leq \varphi \Rightarrow f^c \geq \varphi^c$;
2. $f^{c\bar{c}} \geq f$;
3. $g^{\bar{c}c} \geq g$;
4. $f^{c\bar{c}c} = f^c$.

Démonstration. 1. Par définition.

2. On a :

$$f^{c\bar{c}} = \inf_{y\in\mathcal{Y}} \left(c(x,y) - \underbrace{\inf_{x'\in\mathcal{X}} \left(c(x',y) - f\left(x'\right)\right)}_{\leq c(x,y)-f(x)} \right)_{\geq -(c(x,y)-f(x))}$$
$$\geq \inf_{y\in\mathcal{Y}} \left(c(x,y) - c(x,y) + f(x)\right)$$

3. De même.
4. On a $f^{c\bar{c}} \geq f \Rightarrow f^{c\bar{c}c} \leq f^c$. Avec $g = f^c$, on a $f^{c\bar{c}c} \geq f^c$.

■

4.4 Quelques Cas particuliers

4.4.1 Cas Euclidien Quadratique

On veut ici calculer :

$$\min_{X\sim\alpha, Y\sim\beta} \mathbb{E}\left(\|X-Y\|^2\right) = K - 2 \max_{X\sim\alpha, Y\sim\beta} \mathbb{E}\left(\langle X, Y\rangle\right)$$

où K est une constante ne dépendant que de α et β. On va donc vérifier qu'écrire φ sous la forme $(\alpha, \nabla\varphi\sharp\alpha)$ fonctionne, c'est-à-dire redémontrer le théorème de Brenier :

Preuve du théorème de Brenier. On prend ici $c(x,y) = -\langle x, y\rangle$, qui est symmétrique, on obtient que :

$$f^c(y) = -\sup_x \langle x, y\rangle + f(x) = -(-f)^*(y)$$

pour $\cdot^*$ la transformation de Legendre-Fenchel. Puisqu'on sait que $(-f)^*$ est convexe, f^c est concave.

Si π est optimal pour L_c et f est optimal pour le dual, alors :

$$\mathrm{supp}(\pi) \subseteq \{(x,y) \,|\, f^{cc}(x) + f^c(y) = -\langle x, y\rangle\}$$

En notant $\varphi = -f^{cc}$, φ est convexe et donc :

$$\varphi^*(y) = \sup_x \langle x, y\rangle - \varphi(x) = -f^c(y)$$

On a donc :

$$\mathrm{supp}(\pi) \subseteq \underbrace{\{(x,y) \,|\, \varphi(x) + \varphi^*(y) = \langle x, y\rangle\}}_{\text{Sous-différentielle } \partial\varphi(x)}$$

Si φ est différentiable :

$$\mathrm{supp}(\pi) \subseteq \{(x, \nabla\varphi(x))\}$$

Puisque φ est convexe, φ est différentiable presque partout pour la mesure de Lebesgue, donc si α est absolument continue par rapport à la mesure de Lebesgue, φ est aussi différentiable presque partout pour α. ■

4.4.2 Cas Semi-discret

On suppose ici α absolument continue et $\beta = \sum_{j=1}^m b_j \delta_{y_j}$. On a :

$$\begin{aligned} L_c(\alpha, \beta) &= \max_{f,g \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{Y}), f \oplus g \leq c} \int f \, \mathrm{d}\alpha + \int g \, \mathrm{d}\beta \\ &= \max_{f \oplus g \leq c} \int f \, \mathrm{d}\alpha + \sum_{j=1}^m b_j g(y_j) \end{aligned}$$

Avec $\varphi_v(x) = \min_j c(x, y_j) - v_j$, on peut montrer que :

$$\begin{aligned} L_c(\alpha, \beta) &= \max_{v \in (\mathbb{R}^d)^m} \int \varphi_v \, \mathrm{d}\alpha + \sum_{j=1}^m b_j v_j \\ &= \max_v \int \min_j \left(c\left(x, y_j\right) - v_j \right) \, \mathrm{d}\alpha + \sum_{j=1}^m b_j v_j \end{aligned}$$

Si on considère les cellules de Laguerre :

$$L_j(v) = \{ x \in \mathcal{X} \,|\, \forall j' \neq j, c(x, y_j) - v_j \leq c(x, y_{j'}) - v_{j'} \}$$

On obtient :

$$L_c(\alpha, \beta) = \max_v \sum_j \int_{L_j(v)} \left(c(x, y_j) - v_j \right) \, \mathrm{d}\alpha + \langle b, v \rangle$$

4.4.3 Distance 1-Wasserstein

On se place dans le cas $c(x, y) = d(x, y)$ sur $\mathcal{X} = \mathcal{Y}$.

Proposition 4.5

1. f est c- concave si et seulement si f est δ-lipschitzienne pour $\delta \leq 1$.
2. Is $\mathrm{Lip}(f) \leq 1$, alors $f^c = -f0$

Proposition 4.6 Sous les hypothèses ci-dessus :

$$\mathrm{W}_1(\alpha, \beta) = \max_{\varphi, \mathrm{Lip}(\varphi) \leq 1} \int \varphi \, \mathrm{d}(\alpha - \beta)$$

R Dans le cas discret, $\alpha - \beta = \sum m_k \delta_{z_k}$ avec $\sum m_k = 0$. On a donc :

$$\mathrm{W}_1(\alpha, \beta) = \max_{u_k} \left\{ \sum_k u_k m_k \,\middle|\, \forall k, l, |u_k - u_l| \leq d(z_k, z_l) \right\}$$

R Si on est aussi dans le cas euclidien, la condition lipschitzienne globale peut se remplacer par :

$$\|\nabla \varphi\|_\infty \leq 1$$

5 Transport Optimal Tranché

5.1 Définition et propriétés

Définition 5.1 Given $u \in \mathbb{R}^d$, l'opérateur de tranchage (ou *slicing*) basé sur u est défini par :

$$p_u : x \in \mathbb{R}^d = \langle x, u \rangle$$

On va généraliser cette opération sur les mesures grâce à l'opérateur de poussage $\sharp$: Si $X \sim \mu$, alors $\langle u, X \rangle \sim p_u \sharp \mu$. Dans la suite on note P_u l'application de projection orthogonale sur u.

Définition 5.2 La distance de Wasserstein tranchée par c W_c entre μ et ν est la distance 1-dimensionnelle de Wasserstein basée sur $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ continue est :

$$S\,\mathrm{W}_p(\mu, \nu) = \mathbb{E}_{\theta \sim \mathcal{U}_{\mathbb{S}^{d-1}}}[\mathrm{W}(p_\theta \sharp \mu, p_\theta \sharp \nu)]$$

Démonstration. C'est bien une distance, tkt. ■

Proposition 5.1 Pour toutes deux distributions μ, ν sur $\mathbb{R}^d$:

$$S\,\mathrm{W}_p^p(\mu, \nu) \leq C_{d,p}\,\mathrm{W}_p^p(\mu, \nu)$$

où :

$$C_{d,p} = \frac{1}{d} \int_{\mathbb{S}^{d-1}} \|\theta\|_p^p \, \mathrm{d}\mathcal{U}_{\mathbb{S}^{d-1}}(\theta)$$

5.2 En pratique

On a l'approximation de Monte-Carlo suivante (puisque l'espérance sur la sphère n'est pas calculable en général) :

$$S\,\mathrm{W}_p^p(\mu, \nu) \simeq \frac{1}{K} \sum_{j=1}^{K} W_p(p_{\theta_j} \sharp \mu, p_{\theta_j} \sharp \nu)^p$$

6 Modèles de Flot et Diffusion

Dans cette section, on s'intéresse à l'espace $\mathbb{W}_p$ des mesures de probabilités sur $\mathbb{R}^d$ muni de la distance de p-Wasserstein.

6.1 Transport optimal dynamique

Dans la suite, $c(x, y) = \|x - y\|^2$. On s'intéresse d'abord aux géodésiques sur cet espaces, et à ce qu'elles nous apprennent en apprentissage.

Si on se donne α, β deux mesures de probabilités avec α à densité, et T une application optimale entre α et β, on définit μ_t la distribution de $X_t = (1-t)X_0 + tT(X_0)$ avec $X_0 \sim \alpha$. L'application $t \mapsto \mu_t$ est donc une géodésique dans $\mathbb{W}_2$.

Définition 6.1 On définit $v_t(x) = (T - \mathrm{Id}) \circ T_t^{-1}(x)$, de sorte que $v_t \circ T_t(x) = T(x) - x$ est une constante. C'est le champ de vitesse constant sur les chemins de x à $T(x)$.

On a alors :

$$\int_0^1 \int \|v_t(x)\|^2 \, \mathrm{d}\mu_t \, \mathrm{d}t = \int_0^1 \int \|v_t \circ T_t(x)\|^2 \, \mathrm{d}\alpha \, \mathrm{d}t = \int_0^1 \int \|T(x) - x\|^2 \, \mathrm{d}\alpha \, \mathrm{d}t = \int \|T(x) - x\|^2 \, \mathrm{d}\alpha = \mathrm{W}_2(\alpha, \beta)$$

Définition 6.2 L'ensemble des chemins avec vitesses entre α et β deux mesures de probabilités est :

$$V(\alpha,\beta) = \{(\mu_t, v_t) \,|\, \mu_0 = \alpha, \mu_1 = \beta\}$$

où, pour tout t, $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ et où $v : [0,1]\times\mathbb{R}^d \to \mathbb{R}^d$ est un champ de vitesses et où la paire (μ_t, v_t) vérifie l'équation de continuité (CE).

Proposition 6.1 Si $T = \nabla\varphi$ avec φ convexe, alors T est monotone et pour tout t, T_t est injective.

Définition 6.3 La paire (μ_t, v_t) satisfie l'équation de continuité si :

$$\frac{\mathrm{d}\mu_t}{\mathrm{d}t} + \div(\mu_t v_t) = 0 \tag{CE}$$

où l'égalité est entendue au sens des distributions.

Il reste donc à vérifier que $\mu_t = T_t\sharp\alpha$ et $v_t = (T - \mathrm{Id}) \circ T_t^{-1}(x)$ vérifient l'équation de continuité. On se donne ψ a support compact sur $]0,1[\times\mathbb{R}^d$ et on a :

$$\int_0^1 \int \frac{\mathrm{d}\psi}{\mathrm{d}t}\,\mathrm{d}\mu_t\,\mathrm{d}t = \int_0^1 \int \frac{\mathrm{d}\psi_t}{\mathrm{d}t}(T_t(x))\,\mathrm{d}\alpha\,\mathrm{d}t$$
$$\int_0^1 \int_{\mathbb{R}^d} \langle \nabla_x\psi, v_t(x)\rangle\,\mathrm{d}\mu_t\,\mathrm{d}t = \int_0^1 \int_{\mathbb{R}^d} \langle \nabla_x\psi\circ T_t(x), T(x) - x\rangle\,\mathrm{d}\alpha\,\mathrm{d}t$$

Ainsi :

$$\int_0^1 \int_{\mathbb{R}^d} \left(\frac{\mathrm{d}\psi_t}{\mathrm{d}t}\circ T_t(x) + \langle \nabla_x\psi, v_t(x)\rangle\right)\mathrm{d}\mu_t\,\mathrm{d}t = \int_0^1 \int_{\mathbb{R}^d} \underbrace{\left(\frac{\mathrm{d}\psi_t}{\mathrm{d}t}\circ T_t(x) + \langle \nabla_x\psi\circ T_t(x), T(x) - x\rangle\right)}_{=\frac{\mathrm{d}}{\mathrm{d}t}(\psi(t,T(t,x)))}\mathrm{d}\alpha\,\mathrm{d}t$$

Finalement :

$$\int_0^1 \int_{\mathbb{R}^d} \left(\frac{\mathrm{d}\psi_t}{\mathrm{d}t}\circ T_t(x) + \langle \nabla_x\psi, v_t(x)\rangle\right)\mathrm{d}\mu_t\,\mathrm{d}t = \int_{\mathbb{R}^d} \underbrace{(\psi(1,T(1,x)) - \psi(0,T(0,x)))}_{=0}\,\mathrm{d}\alpha$$

On vérifie donc bien que :

Proposition 6.2

$$W_2^2(\alpha,\beta) \geq \inf_{V(\alpha,\beta)} \int_0^1 \int \|v_t(x)\|^2\,\mathrm{d}\mu_t\,\mathrm{d}t$$

On rappelle que le flot φ de v est défini par $\frac{\mathrm{d}\varphi_t(x)}{\mathrm{d}t} = v(t,\varphi_t(x))$ et $\varphi_0(x) = x$.
Avec la définition de v_t ci-dessus, on voit que T_t est le flot de v_t.

Théorème 6.1 Si v_t est uniformément borné et Lipschitz en x (uniformément en t), si φ_t est son flot et si de plus α est absoluement continue par rapport à la mesure de Lebesgue, alors, $\rho_t = \varphi_t\sharp\alpha$ est l'unique solution de (CE) avec la confition initiale $\rho_0 = \alpha$.

On obtient donc le Théorème suivant :

Théorème 6.2 — Bénamou-Brenier On a :

$$W_2^2(\alpha,\beta) = \min_{V(\alpha,\beta)} \int_0^1 \int \|v_t\|^2 \,\mathrm{d}\mu_t \,\mathrm{d}t$$

la solution étant unique et donnée par l'interpolation de McCann et sa vitesse correspondante v_t.

Démonstration. Utilisant l'équation de la Proposition 6.2 on a déjà une inégalité. En utilisant le Théorème 6.1 ci-dessus, on obtient :

$$\begin{aligned}
\int_0^1 \int_{\mathbb{R}^d} \|v_t(x)\|^2 \,\mathrm{d}\mu_t \,\mathrm{d}t &= \int_0^1 \int_{\mathbb{R}^d} \|v_t \circ \varphi_t(x)\|^2 \,\mathrm{d}\alpha \,\mathrm{d}t \\
&\overset{Fubini}{=} \int_{\mathbb{R}^d} \int_0^1 \|v_t \circ \varphi_t(x)\|^2 \,\mathrm{d}t \,\mathrm{d}\alpha \\
&\overset{Jensen}{\geq} \int_{\mathbb{R}^d} \left\| \int_0^1 \underbrace{v_t \circ \varphi_t(x)}_{\frac{\mathrm{d}\varphi_t}{\mathrm{d}t}(x)} \,\mathrm{d}t \right\|^2 \mathrm{d}\alpha \\
&= \int_{\mathbb{R}^d} \|\varphi(1,x) - x\|^2 \,\mathrm{d}\alpha \\
&\overset{\varphi_1\sharp\alpha=\beta}{\geq} W_2^2(\alpha,\beta)
\end{aligned}$$

Pour l'unicité de la solution, on se donne deux solutions et on introduit leur moyenne, l'inégalité de Cauchy-Schwarz (et son cas d'égalité) nous permet de conclure la preuve. ■

6.2 Couplage par flot

On va appliquer le Théorème 6.1 et construire une paire $(\mu_t, v_t) \sim$ (CE) telle que $\mu_0 = \alpha$, $\mu_1 = \beta$. Le couplage par flot est une manière de construire un poussé en avant entre α et β, mais celui ci ne sera pas nécessairement optimal.

C'est une méthode applicable à la modélisation générative :

1. On échantillonne $\alpha = \mathcal{N}(0, \mathrm{Id})$.
2. On calcule un champ v_θ tel que $\mu_t, v_\theta \sim$ (CE) où μ_t est un chemin de α à β.
3. On intègre v_θ pour optenir φ^θ.
4. On échantillone β en échantillonnant α puis en considérant l'image par φ_1^θ.

Pour définir μ_t, v_t de manière efficace on a l'algorithme suivant :

1. On part d'un couplage $X_0, X_1 \sim \Pi \in \mathcal{U}(\alpha,\beta)$.
2. On définit $X_t = tX_1 + (1-t)X_0$ et μ_t la loi de X_t.
3. On calcule ensuite v_t comme l'espérance de $X_1 - X_0$ sachant $X_t = x$.

Proposition 6.3 La paire (μ_t, v_t) définie ci-dessus vérifie (CE), et donc $(\mu_t, v_t) \in V(\alpha,\beta)$.

7 Barycentres et Lois Multimarginales

8 Barycentre de Sinkhorn

8.1 Régularisation entropique

Considérons deux mesures discrètes $\alpha = \sum_{i=1}^n a_i \delta_{x_i}$ et $\beta = \sum_{j=1}^m b_j \delta_{y_j}$.

Définition 8.1 Le problème de Schrödinger statique s'écrit :

$$\min_{P\in\mathbb{R}_+^{n\times m}, P\mathbf{1}_m=a, P^\top\mathbf{1}_n=b} \langle C,P\rangle + \varepsilon H(P) \tag{Schr}$$

où $H(P)=\sum_{i,j}P_{i,j}\log P_{i,j}$ est l'entropie de Shannon négative de la matrice de transport P, avec la convention $0\log 0=0$. Le paramètre ε est appelé *paramètre de régularisation entropique*, ou *température* en physique statistique.

R H est strictement convexe, puisqu'on se trouve dans le simplexe des matrices à coefficients compris entre 0 et 1.

Proposition 8.1 Si $\varepsilon>0$, le problème de Schrödinger admet une unique solution P_ε. Si $\varepsilon=0$, on retrouve la formulation de Kantorovich, qui peut admettre plusieurs solutions.

Proposition 8.2 Lorsque $\varepsilon\to 0$, la solution du problème de Schrödinger converge vers la solution de Kantorovich qui maximise l'entropie :

$$P_\varepsilon \xrightarrow[\varepsilon\to 0]{} \operatorname{argmin}_P \{H(P) : P \text{ est une solution de Kantorovich}\}$$

À l'inverse, lorsque $\varepsilon\to+\infty$, on retrouve le couplage trivial :

$$P_\varepsilon \xrightarrow[\varepsilon\to+\infty]{} a\otimes b$$

L'intuition derrière cette régularisation entropique est l'ajout d'une fonction de barrière, de manière similaire aux méthodes de points intérieurs en optimisation convexe. Une différence importante étant que l'on utilise ici l'entropie de Shannon négative, et non pas une fonction logarithmique classique. Cela permet de garantir la positivité des coefficients de la matrice de transport P.

Théorème 8.1 Supposons sans perte de généralité que $a_i>0$ pour tout i et $b_j>0$ pour tout j. On a alors :

$$P \text{ est solution du problème de Schrödinger } \iff \begin{cases} P\mathbf{1}_m=a, P^\top\mathbf{1}_n=b \\ \exists u\in\mathbb{R}^n, v\in\mathbb{R}^m, P_{i,j}=u_iK_{i,j}v_j \text{ avec } K_{i,j}=e^{-C_{i,j}/\varepsilon}\end{cases}$$

De manière équivalente, un couplage P est solution s'il existe des vecteurs $u\in\mathbb{R}^n$ et $v\in\mathbb{R}^m$ tels que :

$$P=\operatorname{diag}(u)K\operatorname{diag}(v)$$

où K est la matrice de noyau exponentiel défini par $K_{i,j}=e^{-C_{i,j}/\varepsilon}$.

Démonstration. On veut résoudre le problème de Schrödinger :

$$\min_{P\mathbf{1}_m=a, P^\top\mathbf{1}_n=b} f(P) := \langle C,P\rangle+\varepsilon H(P)$$

Commençons par montrer par l'absurde que si $P^\star$ est solution, alors $P^\star_{i,j}>0$ pour tout (i,j). Supposons qu'il existe (i_0,j_0) tel que $P^\star_{i_0,j_0}=0$. Posons $P^t(1-t)P^\star+t(a\otimes b)$ pour $t\in[0,1]$. Alors P^t est admissible pour tout t. De plus, si l'on pose $g(t)=f(P^t)$, alors $g'(0)=-\infty$ car la dérivée de l'entropie en 0 est infinie. Donc pour t suffisamment petit, $f(P^t)<f(P^\star)$, ce qui contredit le fait que $P^\star$ est solution. Ainsi, $P^\star_{i,j}>0$ pour tout (i,j). Ceci justifie (peu rigoureusement) le fait que l'on peut omettre la contrainte de positivité dans la suite de la preuve.

Dérivons le problème dual de Lagrange. On introduit les multiplicateurs de Lagrange $\{f_i\}_{i=1}^n$ et $\{g_j\}_{j=1}^m$ associés aux contraintes de marges. Le lagrangien s'écrit :

$$L(P,f,g)=\langle C,P\rangle+\varepsilon H(P)+\langle a-P\mathbf{1}_m,f\rangle+\langle b-P^\top\mathbf{1}_n,g\rangle$$

(Ici, les conditions de Slater sont vérifiées car le problème est convexe et admet une solution intérieure, on peut donc échanger le min et le max.)

Remarquons que $\langle a - P\mathbf{1}_m, f\rangle = \langle a, f\rangle - \langle P\mathbf{1}_m, f\rangle = \langle a, f\rangle - \langle P, f\mathbf{1}_m^\top\rangle$. On a alors :

$$\nabla_P L = C + \varepsilon(\log P + 1) - f\mathbf{1}_m^\top - \mathbf{1}_n g^\top.$$

En annulant ce gradient, on obtient :

$$P_{i,j} = e^{(f_i + g_j - C_{i,j})/\varepsilon - 1} = e^{-1} e^{f_i/\varepsilon} e^{-C_{i,j}/\varepsilon} e^{g_j/\varepsilon}.$$

En posant $u_i = e^{f_i/\varepsilon - 1/2}$ et $v_j = e^{g_j/\varepsilon - 1/2}$, on retrouve bien la forme annoncée. ■

8.2 Algorithme de Sinkhorn

On note le produit de Kronecker $\mathrm{diag}(u)z = u \odot z = (u_i z_i)_i$ le produit terme à terme entre les vecteurs u et z. On a donc $P\mathbf{1}_m = u \odot (Kv)$ qui doit valoir a, et de même $P^\top \mathbf{1}_n = v \odot (K^\top u) = b$. On a ainsi le système suivant :

$$\begin{cases} u \odot (Kv) = a \\ v \odot (K^\top u) = b \end{cases} \tag{SchrSys}$$

Théorème 8.2 Le problème de Schrödinger (Schr) est équivalent à la résolution du système (SchrSys).

Proposition 8.3 L'algorithme de Sinkhorn pour la résolution du problème de Schrödinger est le suivant :

- $v \leftarrow \mathbf{1}_m$
- Répéter jusqu'à convergence :
  - $u \leftarrow a/(Kv)$
  - $v \leftarrow b/(K^\top u)$

où l'on note z/w le quotient terme à terme entre les vecteurs z et w.

Cet algorithme est très simple et facilement parallélisable. Chaque itération coûte $\mathcal{O}(n^2)$ opérations, donnant une complexité totale de $\mathcal{O}(Tn^2)$ pour atteindre une précision ε en T étapes ; ceci est à comparer à des algorithmes comme celui du simplexe, qui a une complexité cubique.

Théorème 8.3 Il suffit de $T = \frac{1}{\varepsilon^2}$ itérations pour atteindre une précision ε.

R On a donc une complexité totale de $\mathcal{O}\left(\frac{n^2}{\varepsilon^2}\right)$ pour atteindre une précision ε. Par rapport aux méthodes à points intérieurs de complexité $\mathcal{O}(n^3 \log(\varepsilon))$, la méthode de Sinkhorn est donc plus efficace en termes d'échantillons n, mais moins efficace en termes de précision ε.

Démonstration. S'intéresse à la preuve par contraction qui donne une convergence linéaire, contrairement à une preuve par projection itérative qui donnerait une convergence sous-linéaire, la constante étant meilleure dans le cas linéaire. *Voir les notes de cours pour la preuve.* ■

8.3 Reformulation en divergence de Kullback-Leibler

Définition 8.2 La divergence de Kullback-Leibler entre deux matrices $P, Q \in \mathbb{R}_+^{n\times m}$ est définie par :

$$\mathrm{KL}(P|Q) = \sum_{i,j} P_{i,j} \log\left(\frac{P_{i,j}}{Q_{i,j}}\right) - P_{i,j} + Q_{i,j}$$

Proposition 8.4 $\mathrm{KL}(P|Q) \geq 0$ avec égalité si et seulement si $P = Q$.

L'idée va être d'utiliser $Q = a \otimes b$ comme mesure de référence.

Proposition 8.5 Le problème de Schrödinger (Schr) s'écrit de manière équivalente :

$$\min_{P\in\mathbb{R}_+^{n\times m}, P\mathbf{1}_m = a, P^\top \mathbf{1}_n = b} \langle C, P\rangle + \varepsilon\,\mathrm{KL}(P|a\otimes b).$$

Démonstration. $\mathrm{KL}(P|a\otimes b)$ et $\mathrm{KL}(P|a'\otimes b')$ diffèrent d'une constante additive indépendante de P, donc le minimum est le même. ■

Définition 8.3 Soit $\frac{\pi}{\xi} \in \mathcal{P}(X\times Y)$. Si $\frac{\mathrm{d}\pi}{\mathrm{d}\xi}$ existe, on définit la divergence de Kullback-Leibler entre π et ξ par :

$$\mathrm{KL}(\pi|\xi) = \int_{X\times Y} \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\xi}\right) \mathrm{d}\pi - \pi(X\times Y) + \xi(X\times Y)$$

Si $\frac{\mathrm{d}\pi}{\mathrm{d}\xi}$ n'existe pas, on pose $\mathrm{KL}(\pi|\xi) = +\infty$.

Définition 8.4 Le **problème de Schrödinger général** s'écrit :

$$\inf_{\pi\in\mathcal{P}(X\times Y)} \left\{ \int c\ \mathrm{d}\pi + \varepsilon\ \mathrm{KL}(\pi|a\otimes b) : \pi_1 = \alpha,\ \pi_2 = \beta \right\} \qquad \text{(SchrGen)}$$

R L'information mutuelle entre deux variables aléatoires X et Y de loi jointe π est donnée par $I(X,Y) = \mathrm{KL}(\pi|\pi_1\otimes\pi_2)$. Ainsi, le problème de Schrödinger cherche un couplage π qui minimise le coût total plus une pénalisation de l'information mutuelle entre les deux variables aléatoires. Lorsque $I(X,Y) = 0$, les variables sont indépendantes, et le couplage est donc le produit tensoriel des marges.

Calculons le problème dual de Schrödinger dans ce cadre général. En réutilisant les notations des multiplicateurs de Lagrange, on a :

$$\min_P \max_{f,g} \langle C, P\rangle + \varepsilon\,\mathrm{KL}(P|a\otimes b) + \langle \alpha - P\mathbf{1}, f\rangle + \left\langle \beta - P^\top\mathbf{1}, g\right\rangle$$

Puisque les conditions de Slater sont vérifiées, on peut échanger le min et le max :

$$\max_{f,g} \langle a, f\rangle + \langle b, g\rangle + \min_P \left\langle C - f\mathbf{1}^\top - \mathbf{1}g^\top, P\right\rangle + \varepsilon\,\mathrm{KL}(P|a\otimes b),$$

ce qui est équivalent à :

$$\max_{f,g} \langle a, f\rangle + \langle b, g\rangle - \max_P \left\langle C - f\mathbf{1}^\top - \mathbf{1}g^\top, P\right\rangle - \varepsilon\,\mathrm{KL}(P|a\otimes b).$$

On remarque alors que ceci correspond à la transformation de Legendre-Fenchel de $\mathrm{KL}(\cdot|a\otimes b)$ évaluée en $(f\mathbf{1}^\top + \mathbf{1}g^\top - C)/\varepsilon$. On a donc :

$$\max_{f,g} \langle a, f\rangle + \langle b, g\rangle - \varepsilon\,\mathrm{KL}^\star(z|ab^\top)$$

où $z = (f\mathbf{1}^\top + \mathbf{1}g^\top - C)/\varepsilon$.

Lemme 8.1 La transformation de Legendre-Fenchel de $\mathrm{KL}(\cdot|a\otimes b)$ est donnée par :

$$\mathrm{KL}^\star(Z|Q) = \sum_{i,j} Q_{i,j}\exp(Z_{i,j}) - 1$$

On en déduit le dual de Schrödinger :

$$\max_{f,g} \langle a, f\rangle + \langle b, g\rangle - \varepsilon \sum_{i,j} a_i b_j \exp\left(\frac{f_i + g_j - C_{i,j}}{\varepsilon}\right) + \varepsilon$$

Dans le cas général, pour des mesures non-discrètes, on retrouve une formulation similaire :

$$\inf_{f\in\mathcal{C}(X), g\in\mathcal{C}(Y)} \int f(x)\,\mathrm{d}\alpha(x) + \int g(y)\,\mathrm{d}\beta(y) - \varepsilon \int e^{(f(x)+g(y)-c(x,y))/\varepsilon}\,\mathrm{d}\alpha(x)\,\mathrm{d}\beta(y) + \varepsilon$$

Proposition 8.6 L'**algorithme de Sinkhorn** dans ce cadre s'écrit alors :

- Initialiser g
- Répéter jusqu'à convergence :
  - $f \leftarrow g^{c,\varepsilon} := \mathrm{argmin}_f\, D(f,g)$
  - $g \leftarrow f^{c,\varepsilon} := \mathrm{argmin}_g\, D(f,g)$

où l'on dénote $(\cdot)^{c,\varepsilon}$ l'opérateur de c-transformée réguliarisée par l'entropie, et D la fonctionnelle duale de Schrödinger.

Proposition 8.7 L'**opérateur de c-transformée réguliarisée par l'entropie** s'écrit :

$$f^{c,\varepsilon}(y) = -\varepsilon \log\left(\int e^{(f(x)-c(x,y))/\varepsilon}\,\mathrm{d}\alpha(x)\right)$$

et de manière similaire :

$$g^{c,\varepsilon}(x) = -\varepsilon \log\left(\int e^{(g(y)-c(x,y))/\varepsilon}\,\mathrm{d}\beta(y)\right)$$