Rapport de projet Cassiopée

Réalisé par Mael Spangenberger, Léo Ritchie
et Pierre Chambet

Sujet : Transport optimal pour le filtrage particulaire

Encadrant : Yohan Petetin

Année : 2023-2024

Table des matières

1. Introduction au sujet

Ce projet de groupe s'inscrit dans le cadre du filtrage particulaire, qui est un algorithme ayant pour but d'estimer à partir d'observations données une distribution de probabilités. Il convient donc de d'abord présenter le principe de base avec lequel nous avons travaillé. Nous nous tournons d'abord vers les modèles d'espace. Ils présentent typiquement des équations telles que :

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1})$$
$$\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k, \mathbf{v}_k)$$

où v_k est le vecteur d'état au temps k, u_k est le vecteur de contrôle d'entrée au temps k, w_k est le vecteur du processus de perturbation au temps k, y_k est le vecteur d'observation de sortie au temps k et v_k est le vecteur de mesure du bruit. On ne peut pas directement observer x_k malheureusement, on fait des estimations à partir des y_k. Les w_k et v_k sont inconnus, mais en revanche on connaît parfaitement u_k. Pour faciliter la compréhension, on travaillera ici sur ce modèle :

$$\mathbf{x}_k = A\mathbf{x}_{k-1} + B\mathbf{u}_{k-1} + \mathbf{w}_{k-1}$$
$$\mathbf{y}_k = C\mathbf{x}_k + \mathbf{v}_k$$

On va considérer que w_k est un vecteur aléatoire gaussien de moyenne 0 et de matrice de covariance Q, et v_k l'est également mais avec une matrice de covariance R. Comme w_k est aléatoire pour tout k, x_k le sera aussi. On aura donc des probabilités liées au prochain état : $p(x_k|x_{k-1}, u_k)$.

La fonction de distribution pour la transition d'état est une distribution normale a pour moyenne $Ax_{k-1} + Bu_{k-1}$ et Q pour matrice de covariance. On a donc :

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) =$$
$$= (2\pi)^{-\frac{n}{2}} \det(Q)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_k - A\mathbf{x}_{k-1} - B\mathbf{u}_{k-1})^T Q^{-1} (\mathbf{x}_k - A\mathbf{x}_{k-1} - B\mathbf{u}_{k-1})}$$

De manière similaire, pour y_k on aura $p(y_k|x_k)$ et en utilisant l'équation de départ et les informations sur v_k on a :

$$p(\mathbf{y}_k|\mathbf{x}_k) = (2\pi)^{-\frac{r}{2}} \det(R)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}_k - C\mathbf{x}_k)^T R^{-1}(\mathbf{y}_k - C\mathbf{x}_k)}$$

On travaille ici dans un modèle d'états d'espace markovien, c'est à dire qu'on a l'équation suivante vérifiée :

$$p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \ldots, \mathbf{x}_0) = p(\mathbf{y}_k|\mathbf{x}_k)$$

Formulation du problème :

Le but est d'estimer l'état x_k, y_k à partir des données $x\hat{}_0$ à $x\hat{}_{k-1}$, y_0 à y_k et u_0 à u_{k-1} en général dans ce genre de problèmes. Cependant, les filtres particulaires quant à eux, visent à donner une loi de distribution pour les x_k dépendant des y_k et u_{k-1}.

Pour résoudre ce type de problèmes, on peut donc mettre au point un algorithme du type :

- Génération de nouveaux états x_k basés sur les x_{k-1}
- Calcul des nouveaux poids à l'aide des observations y_k et des poids précédents et normalisation de ceux-ci
- Rééchantillonnage des particules afin d'éviter une dégénérescence des données

Cependant, le souci avec cet algorithme est qu'il ne permet pas de conserver la différentiabilité des particules, ce qui peut s'avérer problématique pour certains cas d'application (comme le tracking de robots par exemple). Fort heureusement, l'article écrit par Adrien Corenflos, James Thornton, George Deligiannidis et Arnaud Doucet propose une méthode afin d'outrepasser cette difficulté.

2. Avancées proposées par l'article

2.1 - Utilisation du transport optimal

L'idée proposée dans l'article "*Differentiable Particle Filtering via Entropy-Regularized Optimal Transport*" est de passer par le principe de transport optimal afin d'améliorer l'étape de reparamétrisation. En minimisant la distance de mesure "*Wasserstein-2*" entre 2 mesures de probabilités, on arrive à obtenir le plan de transport optimal (T.O.). Celui-ci permet d'obtenir les fonctions les plus efficaces pour passer d'une mesure à l'autre.

Le "resampling" à partir du transport optimal repose à la fois sur le poids des particules mais aussi sur leur position. Le plan de T.O. prend alors la forme d'une carte déterministe qui peut être vue comme la transformation d'ensemble (T.E.). Cette approche nécessite de résoudre le problème linéaire vu plus haut à un coût élevé et la T.E. résultante n'est pas différentiable.

2.2 - Régularisation entropique

Pour parer à cela, l'utilisation de T.O. avec régularisation de l'entropie (R.E.) permet de créer une matrice de transport différentiable et moins coûteuse en termes de complexité. On utilise donc une autre version de la distance Wasserstein-2, et en la minimisant on retombe de nouveau sur un plan de transport optimal que l'on peut régulariser.

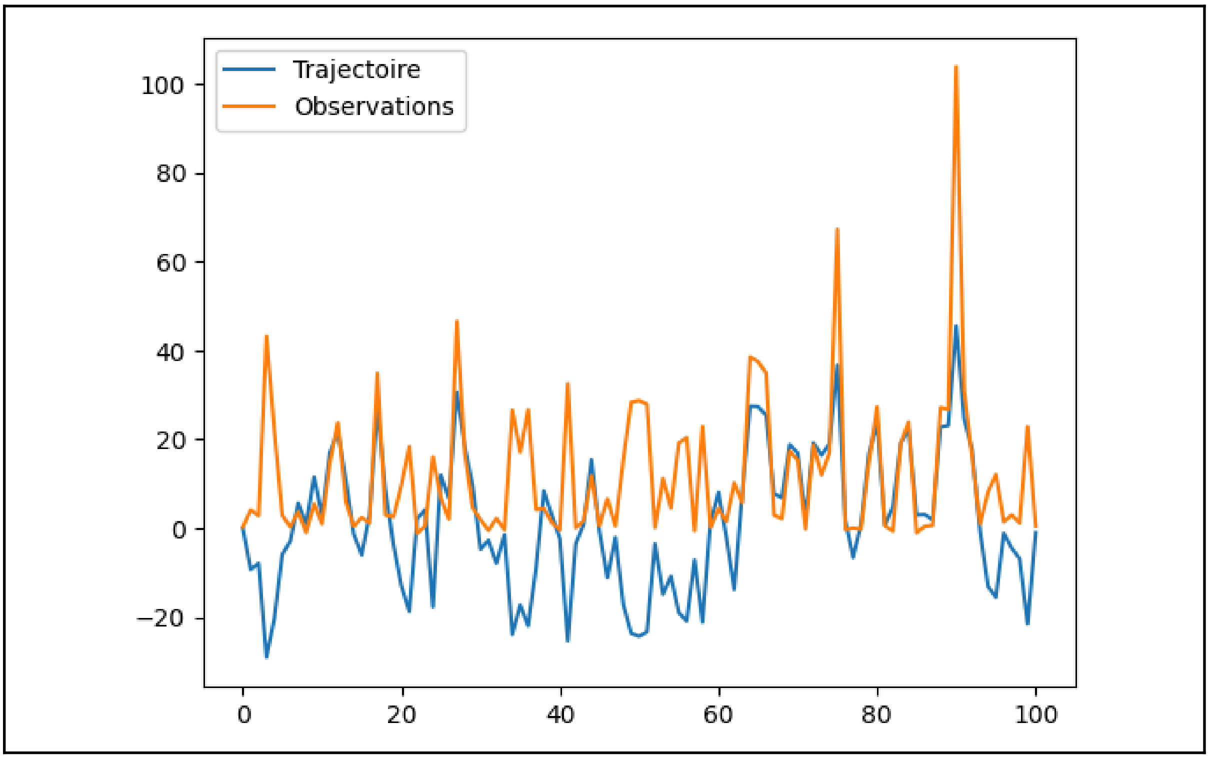
Le problème dual peut quant à lui être maximisé grâce à l'algorithme de Sinkhorn, qui donne des solutions différentiables par rapport aux paramètres d'entrée. Les dérivées du plan de T.O. sont donc accessibles en combinant ces solutions avec la matrice de transport régularisée. En utilisant un nouvel algorithme fourni par l'article avec le T.O. avec R.E., on obtient une T.E. qui est bien différentiable calculée à complexité $O(N^2)$.

Ainsi la technique pour un filtrage particulaire différentiable consiste à utiliser l'algorithme standard pour un filtrage particulaire, mais en utilisant l'astuce de reparamétrisation tout en y implémentant la T.E. différentiable.

3. Implémentation du programme

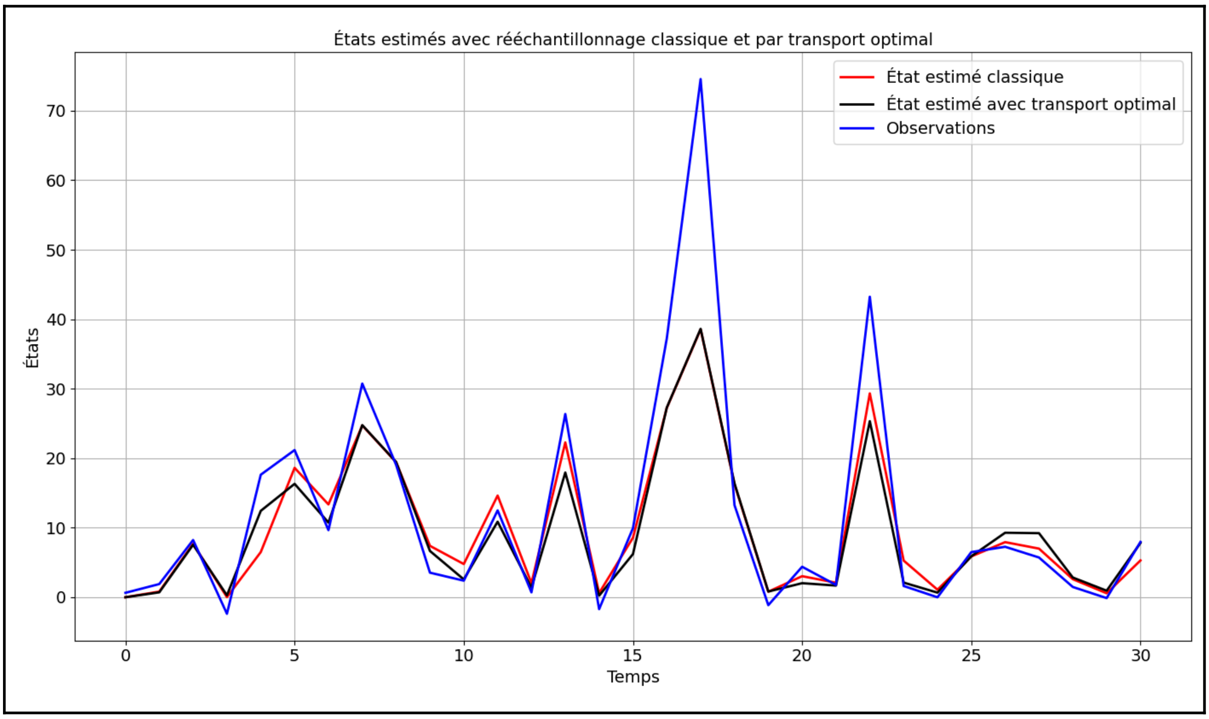
Pour implémenter cet algorithme, nous avons passé quelque temps à comprendre le code fourni dans l'article en question et à en analyser les différentes parties afin de pouvoir le reproduire en version “simplifiée”.

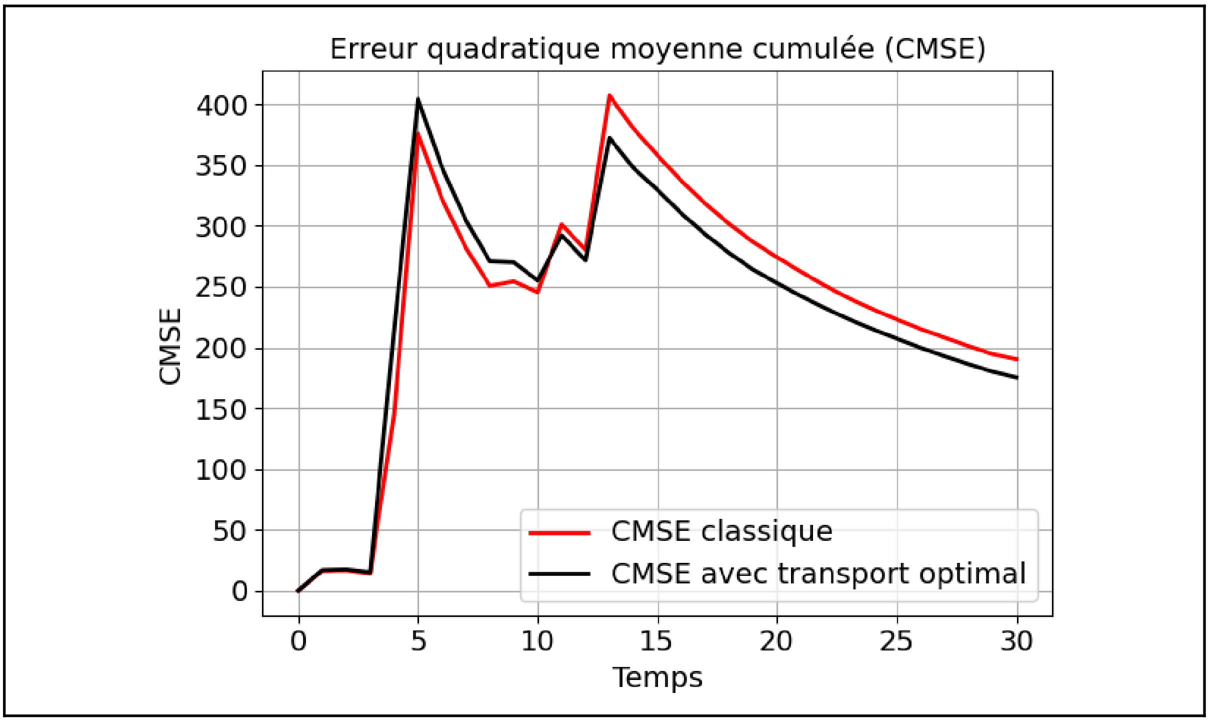
Notre code crée tout d'abord une distribution d'états et d'observations liées à ces états (les x_k et y_k) et les affiche à l'utilisateur. L'utilisateur est bien sûr libre de modifier à sa guise les paramètres donnés en entrée aux lois. Ci-dessous, nous avons une représentation graphique des points obtenus selon les lois.



On fait ensuite tourner 2 algorithmes en parallèle, l'un effectuant le filtrage particulaire avec le rééchantillonnage classique, et l'autre avec le rééchantillonnage basé sur le transport optimal. On représente enfin sur un graphique les courbes obtenues par le biais de ces 2 fonctions et la courbe réelle, afin de permettre une comparaison visuelle des résultats pour l'utilisateur. Pour une comparaison plus analytique, il est bien sûr possible de visualiser les erreurs quadratiques moyennes des 2 méthodes sur un graphique fourni et également sous forme numérique.

Les résultats observés permettent de bien se rendre compte que pour une qualité similaire et un temps de travail proche à basse échelle, l'algorithme amélioré par le transport optimal permet d'obtenir de plus la différentiabilité assurée, ce qui représente un atout majeur, et une EQM légèrement améliorée.





4. Preuves mathématiques

Sommaire

La première preuve est la démonstration de la forme dual du problème du transport optimal. Celle-ci permet de transformer le problème primal en un problème dual plus simple à résoudre.

La seconde preuve est la preuve annexe de l'existence d'un point de selle dans le problème du transport optimal afin de valider une hypothèse de la preuve n°1.

La troisième preuve est la démonstration de la nouvelle forme du problème du transport optimal en ajoutant un terme d'entropie. Ce terme d'entropie permet de rendre le problème plus simple à résoudre. On appelle cela la régularisation entropique.

Preuve 1 : Dualité du problème du transport optimal

Le problème initial du transport optimal (OT) est le suivant :

$$OT = \inf_{P \in \mathcal{S}(a,b)} \left\{ \sum_{i,j} c_{ij} p_{ij} \right\}$$

Ceci est la forme primal du problème. Comme tout problème d'optimisation, il existe une forme dual plus adapté car moins couteûse en puissance de calcul.

La forme dual du problème du transport optimal proposée par l'article est la suivante :

$$OT = \sup_{f,g \in \mathcal{R}(C)} a^t f + b^t g$$

$$\text{où } f = (f_i)\,, g = (g_i)\,, \mathbf{C} = (c_{i,j}) \text{ et } \mathcal{R}(\mathbf{C}) = \{f, g \in \mathbb{R}^N \mid f_i + g_j \leq c_{i,j}, i, j \in [N]\}$$

La preuve consiste à retrouver cette forme dual.

Reprenons le problème initial du transport optimal :

$$OT = \inf_{P \in \mathcal{S}(a,b)} \left\{ \sum_{i,j} c_{ij} p_{ij} \right\}$$

où :

1. **Matrice de coût** $c_{i,j}$:

$$c_{i,j} = \|u_i - v_j\|^2$$

Cela représente la distance euclidienne au carré entre les vecteurs u_i et v_j. Les vecteurs u et v sont les distributions de probabilité à transporter de a à b.

2. **Matrice P** :

$$\mathbf{P} = (p_{i,j})_{i,j \in [N]}$$

C'est une matrice $N \times N$ où chaque élément $p_{i,j}$ se situe dans l'intervalle $[0,1]$. Les composantes de la matrice représentent les probabilités de transport entre les éléments des vecteurs a et b. **P** représente un plan de transport entre deux distributions caractérisées par les vecteurs a et b.

3. **Ensemble** $S(a,b)$:

$$S(a,b) = \left\{ \mathbf{P} \in [0,1]^{N \times N} : \sum_{j=1}^{N} p_{i,j} = a_i, \sum_{i=1}^{N} p_{i,j} = b_j \right\}$$

C'est l'ensemble de toutes les matrices $N \times N$ avec des entrées dans $[0,1]$ telles que :

- La somme de chaque ligne i dans **P** est égale à a_i.
- La somme de chaque colonne j dans **P** est égale à b_j.

L'ensemble $S(a,b)$ définit les plans de transport faisables satisfaisant les contraintes marginales données par a et b.

Début de la preuve :

L'étape clé est l'introduction d'une fonction indicatrice qui permet de reformuler le problème primal en termes de contraintes duales. Ensuite, en appliquant la propriété du point de selle, on peut transformer le problème de minimisation en un problème de maximisation.

D'une part on pose :

$$I(P) = \inf_{P \in \mathcal{S}(a,b)} \left\{ \sum_{i,j} c_{ij} p_{ij} \right\}$$

D'autre part on définit la fonction indicatrice $K(P)$:

$$K(P) = \sup_{f,g \in \mathbb{R}^n} \left\{ \sum_i f_i a_i + \sum_j g_j b_j - \sum_{i,j} (f_i + g_j) p_{ij} \right\}$$

$$K(P) = \begin{cases} 0 & \text{si } P \in \mathcal{S}(a,b) \\ +\infty & \text{sinon} \end{cases}$$

Le problème du transport optimal devient alors :

$$OT = \inf_{P \in \mathcal{S}(a,b)} \{I(P) + k(P)\}$$

Développons son expression :

$$OT = \inf_{P \in \mathcal{S}(a,b)} \left\{ I(P) + \sup_{f,g \in \mathbb{R}^n} \left\{ a^\top f + b^\top y - \sum_{i,j} (f_i + g_j) p_{ij} \right\} \right\}$$

$$= \inf_{P \in \mathcal{S}(a,b)} \left\{ \sup_{f,g \in \mathbb{R}^n} \left\{ a^T f + b^T g - \sum_{i,j} (p_{ij}(f_i + g_j)) + \sum_{i,j} c_{ij} p_{ij} \right\} \right\}$$

Au point de selle de la fonction indicatrice, on a la propriété suivante :

$$\sup_{y \in B} \inf_{x \in A} f(x,y) = \inf_{x \in A} \sup_{y \in B} f(x,y) = f(\bar{x}, \bar{y})$$

$$= \sup_{f,g \in \mathbb{R}^n} \left\{ \inf_{P \in \mathcal{S}(a,b)} \left\{ a^T f + b^T g - \sum_{i,j} (p_{ij}(f_i + g_j)) + \sum_{i,j} c_{ij} p_{ij} \right\} \right\}$$

$$= \sup_{f,g \in \mathbb{R}^n} \left\{ a^T f + b^T g + \inf_{P \in \mathcal{S}(a,b)} \left\{ - \sum_{i,j} (p_{ij}(f_i + g_j)) + \sum_{i,j} c_{ij} p_{ij} \right\} \right\}$$

$$= \sup_{f,g \in \mathbb{R}^n} \left\{ a^T f + b^T g + \inf_{P \in \mathcal{S}(a,b)} \left\{ \sum_{i,j} (c_{ij} - f_i - g_j) p_{ij} \right\} \right\}$$

$\rightarrow$ On pose la nouvelle fonction indicatrice :

$$K'(P) = \inf_{P \in \mathcal{S}(a,b)} \left\{ \sum_{i,j} (c_{ij} - f_i - g_j) p_{ij} \right\} = \begin{cases} 0 & \text{si } (f_i + g_j) \leq c_{ij} \\ -\infty & \text{sinon} \end{cases}$$

En choisissant f et g sont dans un ensemble $R(C)$ défini par la contrainte $f_i + g_j \leq c_{ij}$ qui impose qui impose que $K'(P) = 0$, le problème du transport optimal devient :

$$OT = \sup_{f,g \in \mathcal{R}(C)} a^t f + b^t g$$

$$\text{où } f = (f_i), g = (g_i), \mathbf{C} = (c_{i,j}) \text{ et } \mathcal{R}(\mathbf{C}) = \{f, g \in \mathbb{R}^N \mid f_i + g_j \leq c_{i,j}, i,j \in [1,N]^2\}$$

Nous avons donc retrouvé la forme dual du problème du transport optimal.

Cependant nous avons utilisé une propriété de point de selle pour retrouver le problème dual. Il est donc nécessaire de prouver l'existence d'un point de selle et de valider la démonstration de la propriété utilisée.

Preuve du point de selle : (ou point col)

Soit f une fonction concave-convexe définie sur un produit cartésien $X \times Y$ tel que :

$$\forall (x,y) \in X \times Y, \quad f(x,\bar{y}) \leq f(\bar{x},\bar{y}) \leq f(\bar{x},y)$$

Alors $(\bar{x},\bar{y})$ est appelé "point de selle" et on a la propriété suivante :

$$\max_{y \in Y} \min_{x \in X} f(x,y) = \min_{x \in X} \max_{y \in Y} f(x,y) = f(\bar{x},\bar{y})$$

Prouvons cette propriété.

Montrons en premier l'inegalité suivante :

$$\max_{y \in B} \min_{x \in A} f(x,y) \leq \min_{x \in A} \max_{y \in B} f(x,y)$$

On a déjà que :

$$\min_{x \in A} f(x,y) \leqslant f(x,y) \leqslant \max_{y \in B} f(x,y)$$

Or, pour tout $x \in A$ on a:

$$\max_{y \in B} \min_{x \in A} f(x,y) \leq \max_{y \in B} f(x,y)$$

En prenant le minimum de cette inégalité sur $x \in A$, on obtient :

$$\max_{y \in B} \min_{x \in A} f(x,y) \leq \min_{x \in A} \max_{y \in B} f(x,y)$$

On retrouve ainsi l'inégalité voulue.

En deuxième, montrons que :

$$\min_{x \in A} \max_{y \in B} \leq (\bar{x},\bar{y})$$

Pour cela, il faut revenir à la définition du point de selle, à savoir :

$$\max_{y \in B} f(x,y) = f(\bar{x},\bar{y}) \Longrightarrow f(\bar{x},\bar{y}) \geq \min_{x \in A} \max_{y \in B} f(x,y)$$

De même, on a :

$$\min_{x \in A} f(x,y) = f(\bar{x},\bar{y}) \Longrightarrow f(\bar{x},\bar{y}) \leq \max_{y \in B} \min_{x \in A} f(x,y)$$

En recollant les morceaux on obtient :

$$f(\bar{x},\bar{y}) \leqslant \max_{y \in B} \min_{x \in A} f(x,y) \leqslant \min_{x \in A} \max_{y \in B} f(x,y) \leqslant f(\bar{x},\bar{y})$$

Ainsi, on retrouve la propriété du point de selle :

$$\max_{y \in B} \min_{x \in A} f(x,y) = \min_{x \in A} \max_{y \in B} f(x,y)$$

Preuve 3 : Régularisation entropique du problème

Pour la troisième preuve, il s'agit de démontrer la nouvelle forme du problème du transport optimal en ajoutant un terme d'entropie. On appelle cela la régularisation entropique.

La régularisation entropique ajoute un terme d'entropie à la fonction objectif, rendant le problème plus facile à résoudre numériquement.

La régularisation entropique est une technique fiable car elle garantit l'existence et l'unicité de la solution du problème de transport optimal. De plus, elle permet de contrôler la complexité de la solution en ajustant le paramètre de régularisation.

Cette technique est également viable car les solutions régularisées convergent vers la solution du problème initial lorsque le paramètre de régularisation tend vers 0. Cela garantit que les solutions obtenues sont proches de la solution optimale du problème de transport optimal.

Mise en place :

Posons la distance solution du problème de transport optimal :

$$\mathcal{L}(a,b) = \inf_{P\in S(a,b)} \sum_{i,j} c_{ij} p_{ij} = \inf_{P\in S(a,b)} \langle c, P\rangle$$

avec S(a, b) l'ensemble des couplages admissibles entre les distributions a et b définis précédemment.

On définit alors une nouvelle distance solution du problème de transport optimal régularisé tel que :

$$\mathcal{L}^{\varepsilon}(a,b) = \inf_{P\in S(a,b)} \{\langle c, P\rangle - \varepsilon H(P)\}$$

Avec $\varepsilon > 0$ et $H(P)$ étant l'entropie de la matrice P définie par :

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

Par convention, $H(P) = +\infty$ si $p_{ij} = 0$.

L'idée est donc d'utiliser $-H$ comme fonction de régularisation pour obtenir des approximations de la solution.

Pour assurer l'unicité de la solution, il est nécessaire de vérifier que la fonction de régularisation est convexe.

D'après l'expression de $\mathcal{L}^{\varepsilon}(a,b)$, comme $\langle c, .\rangle$ est convexe, il faut donc que H soit concave, c'est-à-dire que sa matrice hessienne soit négative définie.

Étudions alors la matrice hessienne de H :

$$\frac{\partial H}{\partial p_{ij}} = -\log(p_{ij})$$

$$\frac{\partial^2 H}{\partial p_{ij}^2} = -\text{diag}\left(\frac{1}{p_{ij}}\right)$$

Les coefficients p_{ij} sont compris entre 0 et 1, donc la matrice hessienne de H est négative définie. Ainsi, H est concave, donc la fonction de régularisation est convexe.

Maintenant que nous savons qu'il existe une solution unique, il faut maximiser H pour minimiser $\mathcal{L}^{\varepsilon}(a,b)$ afin de s'en rapprocher.

Il reste également à démontrer que cette solution converge vers la solution du problème de transport optimal initial lorsque ε tend vers 0.

Il faut donc montrer la convergence de $\mathcal{L}^{\varepsilon}(a,b)$ vers $\mathcal{L}(a,b)$ lorsque ε tend vers 0, ie :

$$\lim_{\varepsilon \to 0^+} \mathcal{L}^{\varepsilon}(a,b) = \mathcal{L}(a,b)$$

Pour cela, on pose la matrice $P^{\varepsilon} \in S(a,b)$ solution du problème entropique tel que :

$$P^{\varepsilon} = \arg\min_{P \in S(a,b)} \{\langle c, P \rangle - \varepsilon H(P)\}$$

Dans ce cas,

$$\lim_{\varepsilon \to 0^+} P^{\varepsilon} = \arg\max_{P \in S(a,b)} \{H(P) \mid \langle c, P \rangle = \mathcal{L}(a,b)\}$$

On pose ensuite $\varepsilon_n \to 0$ avec $\varepsilon_n > 0$ pour n $\in \mathbb{N}$.

En utilisant la propriété de compacité de $S(a,b)$, on peut extraire une sous-suite de P^{ε_n} qui converge vers une matrice $P^* \in S(a,b)$ tel que :

$$P^{\varepsilon_n} \to P^*$$

Désormais, on choisit une matrice $P \in S(a,b)$ qui satifasse le problème initial. Autrement dit, P minimise la fonction objectif du problème de transport optimal tel que $\langle c, P \rangle = \mathcal{L}(a,b)$.

On utilise ensuite les conditions d'optimalité de la fonction de régularisation dans l'inégalité suivante :

$$\begin{aligned} 0 &\leq \langle c, P^{\varepsilon_m} \rangle - \langle c, P \rangle \\ 0 &\leq \langle c, P \rangle - \varepsilon_n H(P) + \varepsilon_n H(P^{\varepsilon_n}) - \langle c, P \rangle \\ 0 &\leq \varepsilon_n \left(H(P^{\varepsilon_n}) - H(P) \right) \end{aligned}$$

Comme P^{ε_n} cherche à maximiser l'entropie, on a que H(P^{ε_n}) $\geq$ H(P).

Mais H(P^{ε_n}) est majoré, donc bornée, donc H(P^{ε_n}) $\to$ H(P) lorsque $\varepsilon_n \to 0$.

Ainsi, on a :

$$\lim_{\varepsilon_n \to 0} \left(\langle c, P^{\varepsilon_n} \rangle - \langle c, P \rangle \right) = 0$$

$$\langle c, P^* \rangle = \mathcal{L}(a,b)$$

Autrement dit, P^* est la solution du problème de transport optimal initial et maximise l'entropie parmi tous les plans optimaux de transport.

De plus, la solution P^ε converge vers P^* lorsque ε tend vers 0.

$$\lim_{\varepsilon \to 0} \mathcal{L}^\varepsilon(a,b) = \mathcal{L}(a,b)$$

$$\lim_{\varepsilon \to 0} P^\varepsilon = \arg \max_{P \in S(a,b)} \{H(P) \mid \langle c, P \rangle = \mathcal{L}(a,b)\}$$

Définissons maintenant la distribution de Gibbs appliquée à la matrice de coût c_{ij} est définie par :

$$k_{ij} = \exp\left(-\frac{c_{ij}}{\varepsilon}\right)$$

Le paramètre k_{ij} peut être vu comme une distribution de probabilité entre les points u_i et v_j. Autrement dit, k_{ij} est la probabilité de transporter une unité de masse de a_i à b_j.

L'idée enuite consiste à réduire la distance entre k_{ij} and $S(a,b)$ afin de réduire la distance entre les distributions a et b et donc de trouver le chemin le plus court entre les deux distributions.

Mais comment définir la métrique, la distance entre ces deux distributions ?

On introduit pour cela la divergence de Kullback-Leibler entre les couplages P et K pour définir la distance appelée "entropie relative" :

$$KL(P\|K) = \sum_{i,j} p_{ij} \log\left(\frac{p_{ij}}{k_{ij}}\right) - p_{ij} + k_{ij}$$

Cette mesure n'est pas symétrique, ce qui signifie que $KL(P\|K) \neq KL(K\|P)$. Elle est utilisée pour quantifier la différence entre deux distributions de probabilité. Autrement dit, elle mesure la distance entre les distributions P et K.

Cherchons donc la matrice P qui minimise cette distance.

Reprenons $P^\varepsilon = \arg\min_{P \in S(a,b)} \{\langle c, P \rangle - \epsilon H(P)\}$

Soit une matrice $P \in S(a,b)$ qui satifasse le problème initial :

$$0 \leq \frac{1}{\epsilon}\left[\langle c, P \rangle - \epsilon H(P)\right] - \left[\langle c, P^\varepsilon \rangle - \epsilon H(P^\varepsilon)\right]$$

$$= \sum_{i,j} \left[\frac{c_{ij}P_{ij}}{\epsilon} + P_{ij}\log(P_{ij}) - P_{ij} - \frac{c_{ij}P^\varepsilon_{ij}}{\epsilon} - P^\varepsilon_{ij}\log(P^\varepsilon_{ij}) + P^\varepsilon_{ij}\right]$$

$$= \sum_{i,j} \left[P_{ij}\left(\frac{c_{ij}}{\epsilon} + \log(P_{ij}) - 1\right) - P^\varepsilon_{ij}\left(\frac{c_{ij}}{\epsilon} + \log(P^\varepsilon_{ij}) - 1\right)\right]$$

$$= \sum_{i,j} \left[P_{ij}\log\left(\frac{P_{ij}}{\exp(-c_{ij}/\epsilon)}\right) - P^\varepsilon_{ij}\log\left(\frac{P^\varepsilon_{ij}}{\exp(-c_{ij}/\epsilon)}\right)\right]$$

$$0 \leq \mathrm{KL}(P\|K) - \mathrm{KL}(P^\varepsilon\|K)$$

$$\Rightarrow \mathrm{KL}(P^{\varepsilon}\|K) \leq \mathrm{KL}(P\|K) \quad \forall P \in \mathcal{S}(a,b)$$

Donc P^{ε} est le minimiseur. De pus, on sait que la solution du problème régularisé converge vers la solution du problème initial lorsque le paramètre de régularisation ε tend vers 0 car :

$$\lim_{\varepsilon \to 0} P^{\varepsilon} = P^{*}$$

Pour démontrer maintenant la formulation duale du problème de transport optimal régularisé considérons le problème primal régularisé et introduisons les multiplicateurs de Lagrange afin de dériver la forme duale.

Problème Primal Régularisé par l'Entropie

Nous avons vu que le problème primal de transport optimal régularisé est :

$$\min_{P \in \mathcal{S}(a,b)} \sum_{i,j} c_{ij} p_{ij} - \epsilon H(P)$$

où $H(P) = -\sum_{i,j} p_{ij}(\log p_{ij} - 1)$ etg $\epsilon > 0$.

Formulation du Problème Primal avec les Contraintes de Marges

Le problème primal s'écrit donc :

$$\min_{P \geq 0} \sum_{i,j} c_{ij} p_{ij} - \epsilon \sum_{i,j} p_{ij}(\log p_{ij} - 1)$$

avec les contraintes :

$$\sum_{j} p_{ij} = a_i \quad \forall i$$

$$\sum_{i} p_{ij} = b_j \quad \forall j$$

Introduction des Multiplicateurs de Lagrange

Introduisons les multiplicateurs de Lagrange f et g :

$$L(P,f,g) = \sum_{i,j} c_{ij} p_{ij} - \epsilon \sum_{i,j} p_{ij}(\log p_{ij} - 1) + \sum_{i} f_i \left(\sum_{j} p_{ij} - a_i \right) + \sum_{j} g_j \left(\sum_{i} p_{ij} - b_j \right)$$

Simplification du Lagrangien

Le Lagrangien devient alors :

$$L(P,f,g) = \sum_{i,j} \left[c_{ij} p_{ij} - \epsilon p_{ij}(\log p_{ij} - 1) + f_i p_{ij} + g_j p_{ij} \right] - \sum_{i} f_i a_i - \sum_{j} g_j b_j$$

$$L(P,f,g) = \sum_{i,j} \left[(c_{ij} + f_i + g_j) p_{ij} - \epsilon p_{ij} \log p_{ij} + \epsilon p_{ij} \right] - \sum_{i} f_i a_i - \sum_{j} g_j b_j$$

Calcul du Dual

Pour trouver le dual, nous devons minimiser le Lagrangien par rapport à P. Cela se fait en égalisant à 0 la dérivée partielle de L par rapport à p_{ij}.

$$\frac{\partial L}{\partial p_{ij}} = c_{ij} + f_i + g_j - \epsilon \log p_{ij} = 0$$

$$\epsilon \log p_{ij} = c_{ij} + f_i + g_j$$

$$p_{ij} = \exp\left(\frac{f_i + g_j - c_{ij}}{\epsilon}\right)$$

Ensuite, en substituant p_{ij} dans le Lagrangien, nous obtenons :

$$L(P, f, g) = \sum_{i,j} \left[(c_{ij} + f_i + g_j) \exp\left(\frac{f_i + g_j - c_{ij}}{\epsilon}\right) - \epsilon \exp\left(\frac{f_i + g_j - c_{ij}}{\epsilon}\right) + \epsilon \exp\left(\frac{f_i + g_j - c_{ij}}{\epsilon}\right) \right] - \sum_i f_i a_i - \sum_j g_j b_j$$

Ce qui simplifie à :

$$L(P, f, g) = \sum_{i,j} \epsilon \exp\left(\frac{f_i + g_j - c_{ij}}{\epsilon}\right) - \sum_i f_i a_i - \sum_j g_j b_j$$

En passant à la formulation duale, nous maximisons cette expression par rapport à f et g :

$$\max_{f,g} \left(\sum_i f_i a_i + \sum_j g_j b_j - \epsilon \sum_{i,j} \exp\left(\frac{f_i + g_j - c_{ij}}{\epsilon}\right) \right)$$

Formulation Duale Finale

La formulation duale du problème de transport optimal régularisé par l'entropie est donc :

$$OT_\epsilon = \max_{f,g} \left(\sum_i f_i a_i + \sum_j g_j b_j - \epsilon \sum_{i,j} \exp\left(\frac{f_i + g_j - c_{ij}}{\epsilon}\right) \right)$$

On retrouve bien la forme duale du problème de transport optimal régularisé par l'entropie énoncée dans l'article. Cette formulation montre comment les variables duales f et g jouent un rôle clé dans la détermination de la solution optimale du problème régularisé.

Maintenant que le problème est simplifié, comment calculer et trouver le plan de transport optimal ? C'est là qu'intervient l'algorithme de Sinkhorn, qui permet de résoudre numériquement efficacement le problème de transport optimal régularisé. Cet algorithme consiste à mettre à jour les multiplicateurs de Lagrange f et g pour converger vers la solution optimale.

Implementation de l'Algorithme de Sinkhorn

L'algorithme de Sinkhorn commence avec une matrice de coût K définie par la distribution de Gibbs :

$$K_{ij} = \exp\left(-\frac{c_{ij}}{\epsilon}\right)$$

Ensuite, il procède par les étapes suivantes:

Initialisation :

Initialiser f et g.

Mise à jour itérative :

Alternativement mettre à jour f et g tels que :

$$\mathbf{f_{k+1}} \leftarrow \frac{a}{K\mathbf{g_k}}$$

$$\mathbf{g_{k+1}} \leftarrow \frac{b}{K^T\mathbf{f_k}}$$

Construction de la solution :

La matrice de transport P est obtenue par:

$$P = \text{diag}(f)K\text{diag}(g)$$

Notons ici que P est caractérisé par les vecteurs f et g, c'est à dire avec n + m paramètres (et non n*m comme dans le problème primal). P représente le plan de transport optimal.

Synthèse de notre projet

Ainsi, il est désormais bel et bien de procéder au filtrage particulaire en conservant la différentiabilité des données. Cela nécessite d'implémenter de nouveaux algorithmes basés sur la régularisation par entropie et le transport optimal, mais permet de conserver des temps d'exécution et des résultats similaires à ceux de base.

Ce projet a su nous apporter bien des choses, autant sur un plan académique que sur un plan professionnel. En effet, il nous a fallu savoir gérer à la fois un aspect de travail de groupe et de tâches à accomplir avec le cadre d'un projet de recherche, où l'avancement peut parfois être accéléré ou ralenti, dépendant de notre compréhension du sujet.

Mais grâce à la mise en place d'un planning évolutif, de réunions hebdomadaires avec le tuteur, et de retours continus de sa part afin de pouvoir nous guider dans la bonne direction, nous avons su faire face à ce nouveau challenge, et en ressortons grandis, et prêts à en apprendre toujours plus. Et c'est cela qui fait la marque d'un projet tel que celui de Cassiopée !