

# Optimal Transport/Field Alignment for Data Assimilation - Biblio

Marius Duvillard

14 février 2024

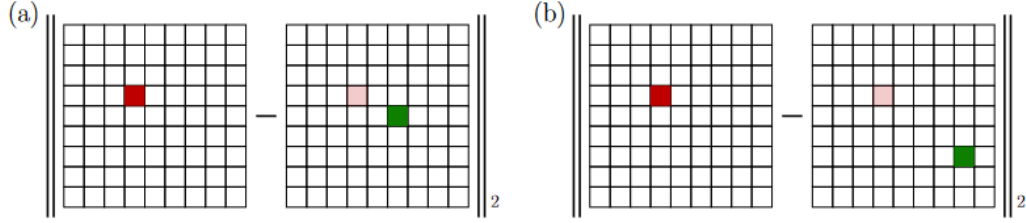
## Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Bocquet, Marc et al., Bridging classical data assimilation and optimal transport, 2023</b>	<b>3</b>
2.1	Proposition . . . . .	5
2.2	Première méthode . . . . .	6
2.3	Deuxième méthode . . . . .	7
<b>3</b>	<b>Ravela et al., Data assimilation by field alignment</b>	<b>7</b>
<b>4</b>	<b>Rosenthal et al., Displacement data assimilation</b>	<b>9</b>
<b>5</b>	<b>Notes personnelles</b>	<b>10</b>

## 1 Introduction

L'assimilation de données consiste à combiner les simulations issues d'un modèle avec les données bruitées issues de l'observation. Cette combinaison nécessite de pouvoir mettre à jour de manière optimale l'état de la simulation en tenant compte par exemple de l'erreur *a priori* du modèle ainsi que l'erreur d'observation. Cette correction est classiquement appliquée sur les intensités des champs de l'état. Cependant, cette correction se trouve faussée par ce que l'on appelle l'effet de *double pénalisation*. Cela se produit lorsque les erreurs du modèle et des données observationnelles sont sur-pénalisées, compromettant l'équilibre nécessaire. Par exemple, un léger déplacement de polluants peut entraîner des prédictions élevées là où aucun polluant n'est observé (voir Figure 1) créant des difficultés d'évaluation du modèle. Cette erreur est répandue dans les géosciences, affectant la prévision météorologique, la chimie atmosphérique, la prévision océanique, etc. L'utilisation de l'erreur quadratique moyenne aggrave ce problème, entravant l'apprentissage efficace des modèles. C'est une composante dominante de l'erreur de représentativité.

La mise à jour de la DA classique peut être limitée à l'espace des valeurs des champs, ce qui entraîne une analyse confinée aux limites d'ébauche et des observations. Ceci devient une lacune majeure lorsque le désaccord entre les observations et l'état de fond résulte d'une erreur de localisation des



**Figure 1.** These two panels schematise the computation of the RMSE of two analysis increments. Those increments are the difference between the truth (left mesh within both Euclidean norm), concentrated here in the red grid-cell and the analysis located in the green grid-cells (right mesh within both Euclidean norm). The increment on panel (a) is the outcome of a better analysis spatially closer to the truth, as compared to the one in panel (b), and yet both increments yield the same RMSE. This verification metrics is hence impacted by the double-penalty error and does not help in discriminating location errors.

FIGURE 1 – Visualisation de l’effet de double pénalisation.

champs ou d’une mauvaise spécification générale de ces champs. La Figure 2 illustre des expériences de DA classique avec des analyses bénéfiques et d’autres inutiles.

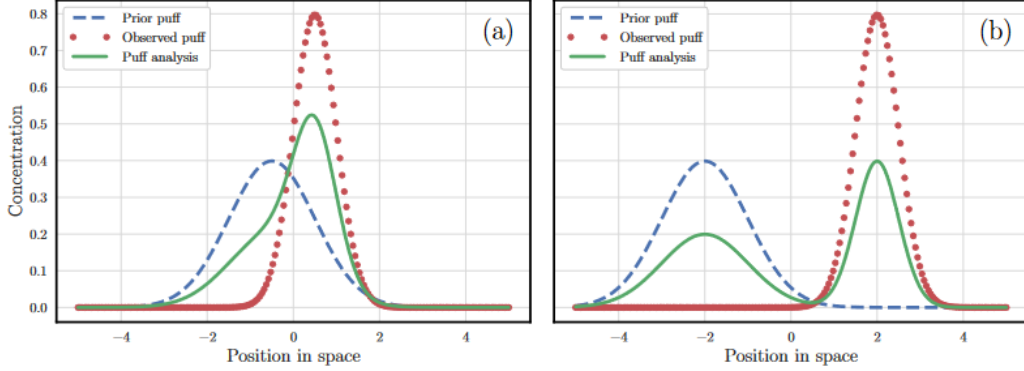
L’objectif des approches qui vont suivre sera de présenter des méthodes d’interpolation non pas dans l’espace euclidien des intensités, mais plutôt dans un espace interpolant qui puisse prendre en compte un déplacement dans l’espace physique.

On peut considérer que l’erreur provient du choix du MSE comme mesure d’erreur, qui est une mesure localisée à une position. Une mesure plus pertinente consiste en la combinaison d’une carte de déplacement suivie de la norme classique MSE (Hoffman et al., 1995; Keil and Craig, 2009). Des métriques basées sur des approches ont pu être proposées. Une approche élégante consiste à appliquer la théorie du *transport optimal* qui est associée à la *distance de Wasserstein*. Des exemples d’utilisation de l’OT peuvent être trouvés chez Farchi et al. (2016) et Vanderbecken et al. (2023). Les travaux de thèse de Feyeux [4] ont permis d’utiliser une distance de Wasserstein pour l’assimilation de données issues d’images. L’article de Bocquet et al. [2] a pour objectif de proposer une approche qui s’applique à des densités ayant des poids potentiellement différents. En effet, Feyeux fait l’hypothèse que les états sont tous dans  $\mathcal{P}(\Omega)$ , états et observations.

Si le transport optimal semble être l’approche la plus générale capable de répondre à cette problématique, on en trouve également d’autres capables de prendre en compte l’erreur de double pénalisation. En particulier, les travaux de Ravela (2007) [8], la thèse de Percival (2008) ou Rosenthal (2018) [9].

Les travaux de Ravela 2007 [8] peuvent trouver une résonance dans une approche OT via une étape d’alignement de champs puis de correction des intensités. Dans la continuité de ces travaux, Ying et al. [10, 11] proposent une approche multi-échelle.

On trouve des formes proches de la distance de Wasserstein en assimilation de données mais dans une toute autre perspective. Il s’agit de transporter la distribution du *prior* vers celle du *posterior* (El Mosehly and Marzouk, 2012; Oliver, 2014; Marzouk et al., 2017; Farchi and Bocquet, 2018; Tamang et al., 2020). On peut l’utiliser aussi pour ajuster les poids du filtre particulaire (Tamang et al., 2021, 2022). Elle a également été utilisée pour comparer des ensembles de forecast (Le Coz et al., 2023; Lledó et al., 2023), ou de précipitation (Duc and Sawada, 2023). Aujourd’hui, on ne traite



**Figure 2.** These two panels illustrate a reasonably beneficial and a probably useless classical DA update. It is assumed that the observations (red dots) and first guess (dashed blue curve), which represent one-dimensional puffs of pollutant, are subject to a location mismatch. With a significant overlap of these fields, panel (a) displays a consistent analysis. By contrast, in panel (b), the analysis state (green curve) fails to propose a state with a significant presence in between the observations and first guess, just as expected since classical DA is based on an interpolation in field values, but neither in space, nor in time.

FIGURE 2 – correction avec l’approche classique à l’aide d’une norme euclidienne et celle escomptée avec une distance de type Wasserstein.

que des cas de pdf de variables aléatoires scalaires. À cause du problème du fléau de la dimension. Dans notre cas, on traite des états qui sont des champs, pas la pdf d’une seule variable. Dans les cas de champs de dimension 2 ou 3.

Un travail de W. Steven Rosenthal [9] trouve des similarités avec notre problématique de vortex avec la recherche d’une fonction de courant pour pouvoir imposer un champ de déplacement incompressible. Ces travaux s’inspirent en partie de la thèse de Percival à l’université de Reading. Tout comme les travaux de Ravela, l’objectif est de corriger l’ébauche avec une transformation qui conserve les volumes. En fait, c’est une condition qui a été soulevée dans Percival 2008 et qui me semble trouver son sens en théorie du réarrangement et dans la théorie dynamique du transport optimal formulée par Bénéamou (voir chapitre 7 de [7] et le chapitre 2 de [4]).

## 2 Bocquet, Marc et al., Bridging classical data assimilation and optimal transport, 2023

Posons quelques notations pour le transport optimal. Le problème consiste à trouver le plan au coût minimal transportant la mesure  $\rho_0$  à  $\rho_1$ . Les mesures sont donc non négatives et ont une intégrale de 1. Un choix d’intégrale peut également être fait. De plus, des généralisations pour des masses différentes ont également été réalisées.

On définit un coût pour tout couple de points  $(x, y) \in \Omega^2$ , généralement fonction de la distance. On présente souvent le coût quadratique

$$\mathcal{C}_{bo}(x, y) = \|x - y\|^2.$$

On introduit l'ensemble des plans de transformations différentiables admissibles  $T$

$$\mathcal{U}_{bo} = \{T : \Omega \mapsto \Omega, \quad \rho_0 = |\partial_x T| \rho_b \circ T\},$$

où  $|\partial_x T|$  est le déterminant de la jacobienne de  $T$  qui prend compte de la déformation de la mesure par la conservation totale de la masse.

On peut alors définir le carré de la distance de *Wasserstein* comme

$$\mathcal{W}_{\mathcal{C}_{bo}}^2(\rho_0, \rho_1) = \min_{T \in \mathcal{U}_{bo}} \int_{\Omega} \mathcal{C}_{bo}(x, T(x)) \rho_0(x) dx,$$

Dont le but est de minimiser le coût total de transport d'une mesure à l'autre. D'un point de vue pratique, il n'est pas possible de résoudre le problème formulé avec un plan de transport déterministe comme énoncé par Monge. Kantorovich introduit une formulation probabiliste au XX<sup>e</sup> siècle. Le plan de transport est remplacé par une mesure de probabilité  $\pi$  définie sur  $\Omega \times \Omega$ .

L'espace des transformations admissibles devient

$$\mathcal{V}_{bo} = \{\pi : \Omega \times \Omega \mapsto \mathbb{R}^+, \quad \rho_o(x) = \int_{\Omega} \pi(x, y) dy, \quad \rho_b(y) = \int_{\Omega} \pi(x, y) dx\}.$$

La distance de Wasserstein au carré devient alors

$$\mathcal{W}_{\mathcal{C}_{bo}}^2(\rho_0, \rho_1) = \min_{T \in \mathcal{V}_{bo}} \int_{\Omega \times \Omega} \mathcal{C}_{bo}(x, y) \pi(x, y) dx dy.$$

Le papier actuel se focalise sur l'utilisation de la distance de Wasserstein lors de la formulation variationnelle de l'analyse.

On part ici de la fonction de coût du 3D-Var (avec la distance Euclidienne)

$$\mathcal{J}_{cl}(\mathbf{x}^a) = \|\mathbf{y}^b - \mathbf{x}^a\|_2^2 + \|\mathbf{y}^o - \mathbf{H}\mathbf{x}^a\|_2^2,$$

où  $\mathbf{y}^b$  est la prédiction du forecast, où  $\mathbf{y}^o$  est le vecteur des observations. En substituant la norme Euclidienne par la norme de Wasserstein on obtient la fonction de coût suivante

$$\mathcal{J}_w(\mathbf{x}^a) = \mathcal{W}_2^2(\mathbf{y}^b, \mathbf{x}^a) + \mathcal{W}_2^2(\mathbf{y}^o, \mathbf{H}\mathbf{x}^a).$$

Il est nécessaire d'équilibrer les deux distances. On appelle l'état d'analyse le *barycentre de Wasserstein*.

Il remarque dans les travaux de Feyeux [4] que cette formulation peut être inconsistente. Dans le cas où l'état est complètement observé ( $\mathbf{H}$  est l'identité), l'assimilation se déroule sans encombre. La principale raison à cela est de supposer que les densités d'origine et objectif ont la même masse. Ce qui s'applique pour les deux termes de calcul de distance de Wasserstein, à la fois que  $m(\mathbf{x}^a) = m(\mathbf{y}^b)$  et  $m(\mathbf{H}\mathbf{x}^a) = m(\mathbf{y}^o)$ . En supposant finalement que  $m(\mathbf{y}^b) = m(\mathbf{y}^o)$ , l'égalité  $m(\mathbf{x}^a) = m(\mathbf{H}\mathbf{x}^a)$  est difficilement atteignable pour  $\mathbf{x}^a \in \mathbb{R}^{N_a} \setminus \ker(\mathbf{H})$ . Il faut pour cela proposer une méthode qui puissent s'accommoder d'avoir des densités de masses différentes. Une résolution a été proposée dans la communauté statistique par Chizat et al. (2018). Ici une solution orientée DA.

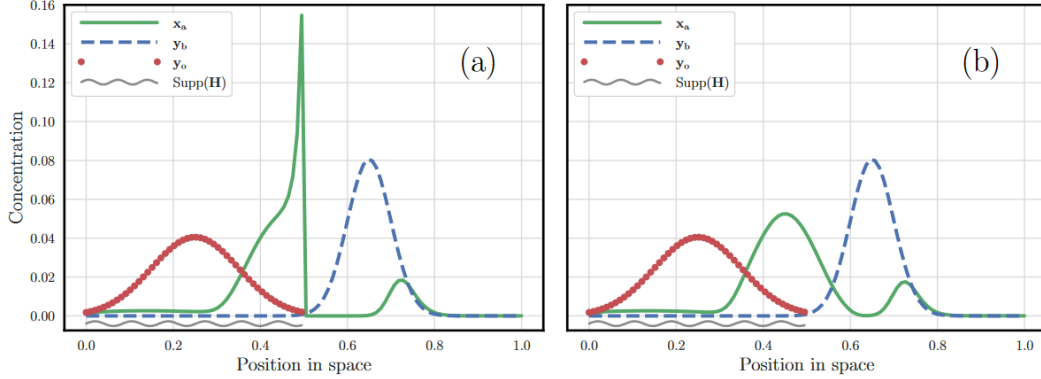


FIGURE 3 – Cas où le support n'est pas défini sur l'ensemble du domaine. Afin de conserver la masse, la solution va créer un pic d'intensité à la frontière. On souhaite un comportement comme à droite.

## 2.1 Proposition

Hypothèses : champs physiques non-négatifs, mais pas nécessairement l'ébauche et l'observation et  $H$  est supposé linéaire.

Ils ont introduit l'équation (18) dans leur papier qui est une nouvelle fonction coût

$$L_w(\mathbf{x}^a) = \min_{\mathbf{x}^b \in \mathcal{O}_{N_b}^+, \mathbf{x}^o \in \mathcal{O}_{N_o}^+} \{ \zeta_b(\mathbf{y}^b - \mathbf{x}^b) + \zeta_o(\mathbf{y}^o - \mathbf{H}\mathbf{x}^o) + W_{C_{ba}}(\mathbf{x}^b, \mathbf{x}^a) + W_{C_{oa}}(\mathbf{x}^o, \mathbf{x}^a) \}.$$

Elle introduit en fait deux nouvelles variables d'optimisation  $\mathbf{x}^b$  et  $\mathbf{x}^o$ . L'objectif est de pouvoir équilibrer toutes les densités. Cette fonction doit être minimisée en fonction de  $\mathbf{x}^b, \mathbf{x}^a, \mathbf{x}^o$ , ses trois quantités étant libres. Pour cela, les auteurs présentent la forme duale du problème. Cette formulation duale du problème complet est assez lourde et nécessite de déterminer un plan de transport entre les trois densités. En introduisant des multiplicateurs de Lagrange, en faisant permuter les variables à minimiser ou maximiser, on obtient finalement un problème à deux variables duales à minimiser sur un polyèdre.

$$\mathcal{L}^* = \max_{(\mathbf{f}_b, \mathbf{f}_o) \in \mathcal{U}_{bo}(\mathbf{C}_{ba}, \mathbf{C}_{ba}, \mathbf{H})} \{ \mathbf{f}_b^T \mathbf{y}^b + \mathbf{f}_o^T \mathbf{y}^o - \|\mathbf{f}_b\|_B^2 - \|\mathbf{f}_b\|_R^2 \}$$

où  $\mathcal{U}_{bo}(\mathbf{C}_{ba}, \mathbf{C}_{ba}, \mathbf{H})$  est l'espace admissible tel que

$$\mathcal{U}_{bo}(\mathbf{C}_{ba}, \mathbf{C}_{ba}, \mathbf{H}) = \left\{ \mathbf{f}_b, \mathbf{f}_o : \forall i, j, k, \quad f_b^i + f_o^l H_l^j \leq C_{ba}^{ik} + C_{oa}^j k \right\}$$

Il présente qq simplification mais on comprends que ce problème doit être résolu en plusieurs étapes. Pour cela, il propose deux méthodes. Une première qui va permettre d'estimer le plan  $\mathbf{P}^{bo}$  à l'aide d'une régularisation entropique (via la divergence  $\mathcal{KL}$ ). La solution assimilée est ensuite le barycentre, c'est à dire l'interpolant au sens de McCann. Ainsi  $\mathbf{x}^a$  est un interpolant sur la géodésique entre  $\mathbf{x}^o$  et  $\mathbf{x}^b$  dans un espace de Riemann.

Il propose une autre formulation plus directe par la suite. (Note : Je ne pense pas cette dernière soit la plus applicable à notre cas de simulation sans maillage).

La résolution se base sur les méthodes dites *Smooth Optimal Transport* et des développement des termes de régularisations duaux comme dans [1].

## 2.2 Première méthode

Pour la résolution, il introduit une régularisation entropique avec la divergence  $\mathcal{KL}$ . Cette régularisation permet de convexifier le problème et permet surtout d'avoir des densités  $\mathbf{x}^b, \mathbf{x}^o$  de masse différente de 1. A noté que  $\mathbf{x}^b, \mathbf{x}^o$  seront de même masse.

On rappelle que la définition de  $\mathcal{KL}$  pour deux distributions quelconque  $\mathbf{p}, \mathbf{q}$

$$\mathcal{KL}(\mathbf{p}||\mathbf{q}) = \sum_i p_i - \ln \frac{p_i}{q_i} - p_i - q_i.$$

Le problème primal se simplifie et devient

$$L = \min_{\mathbf{x}^b \in \mathcal{O}_{N_b}^+, \mathbf{x}^o \in \mathcal{O}_{N_o}^+} \{ \zeta_b(\mathbf{y}^b - \mathbf{x}^b) + \zeta_o(\mathbf{y}^o - \mathbf{H}\mathbf{x}^o) + W_{C_{bo}}(\mathbf{x}^b, \mathbf{x}^o) \}.$$

En ajoutant le terme de régularisation

$$L_\varepsilon(\mathbf{x}^a) = \min_{\mathbf{x}^b \in \mathcal{O}_{N_b}^+, \mathbf{x}^o \in \mathcal{O}_{N_o}^+} \{ \zeta_b(\mathbf{y}^b - \mathbf{x}^b) + \zeta_o(\mathbf{y}^o - \mathbf{H}\mathbf{x}^o) + \min_{\mathbf{P} \in \mathcal{U}_{bo}} (\text{Tr}(\mathbf{C}_{bo}\mathbf{P}^{bo}) + \varepsilon \mathcal{KL}(\mathcal{P}||\boldsymbol{\nu})) \}.$$

Le problème dépend donc de deux paramètres supplémentaires  $\varepsilon > 0$  et  $\boldsymbol{\nu}$ .  $\mathcal{U}_{bo}$  est l'espace admissible des plans de transfert entre  $\mathbf{x}^b$  et  $\mathbf{x}^o$ .  $\boldsymbol{\nu}$  est un plan de transport a priori. Or il ne peut pas être choisi à partir de  $\mathcal{U}_{bo}$  celui-ci n'étant pas connu. Ainsi ils ont décidé de le choisir constant en utilisant un prior indépendant  $\mathbf{x}^b$  et  $\mathbf{x}^o$ .

On peut retrouver la formule dual en équation (30) du papier

$$\begin{aligned} \mathcal{L}_\varepsilon = & \max_{\mathbf{f}_b \in \mathbf{R}^{N_b}, \mathbf{f}_o \in \mathbf{R}^{N_o}} \mathbf{f}_b^T \mathbf{y}^b + \mathbf{f}_o^T \mathbf{y}^o - \zeta_b^*(\mathbf{f}_b) + \zeta_o^*(\mathbf{f}_o) + \\ & \min_{\mathbf{P} \in \mathcal{O}_{b,o}^+} \left( \varepsilon \sum_{ij} \left\{ P_{ij} \ln \frac{P_{ij}}{\nu_{ij}} - P_{ij} + \nu_{ij} \right\} + P_{ij} \left\{ C_{bo}^{ij} - f_b^i 1_o^j - H_l^j f_o^l 1_b^i \right\} \right) \end{aligned}$$

Finalement, par optimisation, on peut obtenir deux variables duales  $\mathbf{f}_b, \mathbf{f}_o$  qui permettent d'obtenir le plan  $\mathbf{P}$ .

Ainsi, en récupérant les marginales, on obtient

$$\mathbf{x}^b = \mathbf{P} \mathbf{1}_o \quad \mathbf{x}^o = \mathbf{P}^T \mathbf{1}_b.$$

L'étape suivante est finalement de trouver le barycentre au sens de la distance de Wasserstein, ce qui signifie résoudre

$$J = \min_{\mathbf{x}^a \in \mathcal{O}_{N_a}^+} \{ W_{C_{ba}}(\mathbf{x}^b, \mathbf{x}^a) + W_{C_{oa}}(\mathbf{x}^o, \mathbf{x}^a) \}.$$

Et cette étape offre de nombreuses variantes dans la littérature. Par exemple, l'article de Cuturi [3] offre des développements pour calculer le barycentre sur un espace de distribution empirique en faisant varier alternativement le support et les intensités des atomes. Les travaux de Feyeux [4] offrent également de nombreuses manières de différencier et de minimiser cette fonctionnelle. En particulier, il présente les différentes dérivées pour la distance de Wasserstein. Finalement, on se place dans le cadre d'une fonctionnelle à minimiser suivant un espace de paramètres donné et on pourra s'inspirer du chapitre 7 de [7]. Ainsi, dans ce chapitre, Peyré conseille aussi d'utiliser la différentiation automatique.

## 2.3 Deuxième méthode

Pour résoudre le problème, Boucquet propose finalement une version simplifiée. On ne se base que sur deux transferts différents  $\mathbf{P}^{ba}, \mathbf{P}^{oa}$ . On régularise pour obtenir

$$\begin{aligned} \min_{\mathbf{x}^b \in \mathcal{O}_{N_b}^+, \mathbf{x}^o \in \mathcal{O}_{N_o}^+, \mathbf{x}^a \in \mathcal{O}_{N_a}^+} & [\zeta_b(\mathbf{y}^b - \mathbf{x}^b) + \zeta_o(\mathbf{y}^o - \mathbf{H}\mathbf{x}^o) + W_{C_{ba}}(\mathbf{x}^b, \mathbf{x}^a) + W_{C_{oa}}(\mathbf{x}^o, \mathbf{x}^a) \\ & + \min_{\mathbf{P}^{ba} \in \mathcal{U}_{ba}, \mathbf{P}^{oa} \in \mathcal{U}_{oa}} \{\varepsilon \mathcal{K}(\mathbf{P}^{oa} | \mathbf{v}^{oa}) + \varepsilon \mathcal{K}(\mathbf{P}^{ba} | \mathbf{v}^{ba}) + \text{Tr}(\mathbf{C}_{oa}^T \mathbf{P}^{oa}) + \text{Tr}(\mathbf{C}_{ba}^T \mathbf{P}^{ba})\}]. \end{aligned}$$

La résolution va directement prendre en compte le fait que  $x_a$  est le Barycentre et par minimisation selon une variable, on retrouve une fonctionnelle à minimiser avec seulement deux variable duales à partir desquelles on peut retrouver  $x_a$ .

## 3 Ravela et al., Data assimilation by field alignment

Une méthode pour tenir compte des erreurs de position en appliquant un déplacement à imposer. La méthode est réalisée en deux étapes. D'abord, la détermination d'un champ de déplacement, puis la correction des intensités. L'avantage de cette méthode est qu'elle est adaptable à de nombreux filtres, en particulier les filtres d'ensemble. En effet, la seconde étape s'apparente à un filtre classique. Il présente tout d'abord en quoi la formulation classique dans un problème de position entraîne une inflation de la matrice de covariance. La matrice issue de la perturbation d'une translation  $\lambda$  devient :  $C_\lambda = C_{XX} + \sigma_\lambda^2 C_{\Delta\Delta}$  où  $\Delta$  est la déviation du gradient (8).

Pour réaliser l'alignement, on part de la formule de Bayes comme pour les filtres d'intensité. Cependant, on introduit une nouvelle variable spatiale  $\mathbf{q}$  vecteur de déplacement. Dans leur travail l'état est défini sur une grille dont les positions sont  $\mathbf{r}$ . L'état est donc défini par les valeurs du champ  $X = X(\mathbf{r})$ .

On défini  $\mathbf{q}$  sur la même grille tel que  $X(\mathbf{r} - \mathbf{q})$  est le déplacement de X par  $\mathbf{q}$ .

On réécrit la formule de Bayes en introduisant cette nouvelle variable

$$P(X, \mathbf{q} | Y) \propto P(Y | X, \mathbf{q}) P(X^f | \mathbf{q}) P(\mathbf{q})$$

La Likelihood  $P(Y | X, \mathbf{q})$  est similaire à la likelihood sur les intensités mais en appliquant au préalable le déplacement de telle sorte que les observations sont conditionnées par  $X(\mathbf{r} - \mathbf{q})$ .

On suppose donc que  $Y = H(X(\mathbf{r} - \mathbf{q})) + \boldsymbol{\eta} \sim \mathcal{N}(0, R)$ , où les observations restent bien fixes.

De même le prior sur l'intensité est inchangé en considérant au préalable la transformation  $\mathbf{q}$ . Mais il faut remarquer que la matrice de covariance est dépendante de  $\mathbf{q}$ . Dans un cas de filtre d'ensemble il s'agit de calculer la matrice également après transformation.

Finalement, le prior sur le déplacement  $P(\mathbf{q})$  est un élément additionnel. Pour cela, on introduit une fonction d'énergie qui va régulariser les déplacements. C'est elle qui introduit un coût de déplacement. On peut penser au champ de déplacement comme à un champ d'écoulement lissé. Cette propriété de lissage amène donc à considérer des pénalisations de Tikhonov qui vont pénaliser à la fois le gradient et la divergence. Mais on peut tout à fait modifier notre a priori sur la distribution de  $\mathbf{q}$ . La prendre uniforme n'apporterait pas d'information, la prendre gaussienne demande de définir une manière appropriée pour la définir. La contrainte ici reste locale, mais elle introduit une forte régularité par pénalisation.

En prenant l'opposée de la log likelihood on obtient la fonction de coût associée  $J(X, \mathbf{q})$ . On remarquera que un terme inhabituel lié à la dépendance de  $P$  à  $\mathbf{q}$ . On note par la suite  $\mathbf{p} = \mathbf{r} - \mathbf{q}$  où  $\mathbf{r}$  est la coordonnées dans l'espace non déformé.

Ainsi on obtient

$$J_2(X, \mathbf{q}) = \frac{1}{2} \|(X - X^f)(\mathbf{r} - \mathbf{q})\|_{P(\mathbf{q})}^2 + \frac{1}{2} \|Y - H(X(\mathbf{r} - \mathbf{q}))\|_R^2 + L(\mathbf{q}) + \frac{1}{2} \ln |B(\mathbf{q})|.$$

Pour résoudre le problème on a besoin de  $P(\mathbf{q})$  et dériver toute la fonction coût. On peut utiliser pour cela un ensemble comme en EnKF pour l'estimer. Pour cela on suppose connu les déplacements associés à chacun des membres  $\mathbf{q}_s$ . On note  $\mathbf{p}_s = \mathbf{r}_s - \mathbf{q}_s$ . et on estime donc  $B_Q$ . Finalement, on finit par avoir une version Hybrid de la fonction de coût.

Il choisit également de fixer les déplacement dans l'évaluation de  $B_Q$  pour éviter de devoir les différentier et modifier dans la version itérative. Il peut ainsi définir les gradients en fonction de  $\mathbf{q}_s$ . Une autre proposition consiste à séparer la phase d'alignement et de correction des intensités, c'est la méthode séquentielle. Pour cela, à partir de la fonction de cout, on écrit les équation d'Euler-Lagrange. On écrit la fonction de coût différentiée selon  $\mathbf{q}$  puis selon  $\mathbf{X}$ . Je note implicitement  $s$  en indice car cela est réalisé parallèlement sur l'ensemble.

$$\begin{aligned} \frac{\partial J_s}{\partial \mathbf{q}} &= (\nabla X|_{\mathbf{p}} - \nabla X^f|_{\mathbf{p}})^T B_Q^{-1}(X(\mathbf{p}) - X^f(\mathbf{p})) + \nabla X|_{\mathbf{p}} H^T R^{-1}(HX(\mathbf{p}) - Y) + \frac{\partial L}{\partial \mathbf{q}} = 0 \\ \frac{\partial J_s}{\partial \mathbf{X}} &= B_Q^{-1}(X(\mathbf{p}) - X^f(\mathbf{p})) + H^T R^{-1}(HX(\mathbf{p}) - Y) = 0. \end{aligned}$$

On commence par fixer  $X$  à  $X^f$  dans la première équation, ainsi que  $Q$  ce qui permet d'obtenir

$$\frac{\partial L}{\partial \mathbf{q}} + \nabla X|_{\mathbf{p}} H^T R^{-1}(HX(\mathbf{p}) - Y) = 0$$

Le choix de pénaliser la divergence de  $\mathbf{q}$  permet d'obtenir une équation de Poisson à résoudre avec un terme source non linéaire. Il propose de résoudre cela avec une méthode de type BFGS. Après des essais en utilisant l'autodifférentiation, je constate que ce sont des problèmes fortement non linéaires. Dans la littérature, une approche multiéchelle de cette méthode a été proposée. De cette manière, le déplacement appliqué est une composition de modes avec plus ou moins de raffinement. Selon moi, cela devrait améliorer la convergence.

Finalement, la deuxième étape consiste à résoudre la seconde équation avec  $\mathbf{q}$  fixé. On peut réévaluer  $B_Q$ . On remarque que l'on obtient la même mise à jour qu'avec un filtre de Kalman. Avec la formule de Sherman-Morrison-Woodbury, on obtient la correction de l'intensité  $\mathbf{X}$  comme combinaison linéaire du forecast et de l'innovation.



## 4 Rosenthal et al., Displacement data assimilation

Dans cet article les auteurs décide de corriger l'ébauche. Il parte de l'évaluation de l'erreur classique pour un champ de vorticit    $\omega$  et il cherche  $\omega_a$  qui minimise

Ainsi, il part de l'id  e de r  soudre 2 fonctions co  t    la suite, une pour corriger la position (champ de d  placement param  tr   par  $a$ ) et l'autre qui va corriger le l'  bauche qui aura   t   d  plac  . On retrouve ainsi la m  me d  marche que Ravela.

$$\begin{aligned}\mathcal{J}_p(a) &= \frac{1}{2}\|d - h_a(a)\|_{R^{-1}}^2 + \frac{1}{2}\|a - a^f\|_{P_a^{-1}}, \\ \mathcal{J}_a(\omega) &= \frac{1}{2}\|d - h(\omega)\|_{R^{-1}}^2 + \frac{1}{2}\|\omega - \omega^f\|_{P_\omega^{-1}},\end{aligned}$$

avec  $d$  les observations et  $R^{-1}$  la matrice de covariance sur les observations qui sont utilis  es plusieurs fois. Les op  rateurs d'observations sont diff  rents (uniquement d  terminant pour la diff  rentiation), ainsi que les statistiques sur les d  placements  $a^f$  et  $P_a$ .

La premi  re chose qui est r  alis  e est de d  finir une transformation. Il d  finit pour cela un map-page entre les variables dans l'espace d  form    $Z$  et dans l'espace non d  form    $z$  (j'ai l'impression que la convention est invers  e...). Les coordonn  es  $Z$  permettent ensuite d'  valuer l'erreur avec la ground truth ou d'  valuer la pr  diction des observations. Par exemple, pour le cas o   le champ est directement observ   sur une discr  tisation eul  rienne.

$$J_a(\omega) = \|\omega^t(Z) - \omega(Z)\|^2$$

En d  finissant une transformation inversible  $Z = \Phi(z; a)$ , on peut ensuite appliquer la transformation pour obtenir le champ d  plac   de  $\omega^f$  d  fini avec  $z$

$$J_p(a) = \frac{1}{2}\|\omega^t(Z) - \omega^f \circ \Phi^{-1}(Z; a)\|^2.$$

Quel est donc cette transformation ? Elle doit respecter la conservation de la vorticit   totale. Plut  t qu'une transformation canonique qui    le probl  me d'  tre implicite. Ici, il propose d'int  grer la fonction de courant sur un interval de temps fixe. Ainsi  $\Phi$  devient

$$\Phi(x, y; a) = (x, y) + \int_0^1 (-\Psi_y[x(t), y(t); a], \Psi_x[x(t), y(t); a]) dt.$$

On choisie alors de param  trer la fonction courant  $\Psi$  lin  airement avec  $a$  par exemple  $\Psi(x, y; a) = Ba$ . L'inverse de la transformation est alors obtenue en substituant les coordonn  es et en prenant  $a \mapsto -a$ . Comme cette hypoth  se se base sur de petite transformation, on pourra donc trouver des s  quence de  $(a_{t_i})_{i=1}^n$  pour pouvoir faire l'int  gration.

En ce qui concerne la r  gularisation de  $a$ , il propose simplement de faire de la propagation de  $P_\omega$ . Il suffit pour cela de connaitre  $T_{\omega a}$  ainsi que son (pseudo)inverse  $T_{a\omega}$  ce qui donne  $P_a = T_{a\omega}P_\omega T_{\omega a}$ . Dans le cas o    $P_\omega$  n'est pas connu, on peut faire de la r  gularisation sur la d  formation. Toutefois, il signale que  $P_a$  peut ne plus   tre SDP. Ainsi, une   tape de projection est n  cessaire.

Mais, l'utilisation d'un ensemble permet de r  gler ce souci. En effet, Pour un ensemble  $X^f$  de forecast, on peut approximer  $P_a$  comme

$$P_a = \frac{1}{N-1} [T_{a\omega}(X - \bar{x}\mathbf{1}\mathbf{1}^T)]^T [T_{a\omega}(X - \bar{x}\mathbf{1}\mathbf{1}^T)]$$

En utilisant l'approximation tangentiel de l'opérateur  $h_p$  qui relie les paramètres  $a$  de déformation et les prédictions, on peut obtenir alors la correction sur les positions  $A = a_{j=1}^{N_{ens}}$

$$A = L [d_V - h_a(X^f)],$$

où  $L$  est défini comme  $L = P_a H^T (R + H_p P_a H_p^T)$  avec  $H_p$  la linéarisation de l'opérateur d'observation en fonction de  $a$  autour de  $0^1$

$$\begin{aligned} H_p &= \nabla_a h_p(A; X^f) \\ &= \nabla_a h_a(X^f \circ \Phi^{-1}(Z; A)) \\ &= \nabla_\omega h_a(X \circ \Phi^{-1}) \cdot \nabla_a [X \circ \Phi^{-1}(Z; A)] \\ &= \nabla_\omega h_a(X \circ \Phi^{-1}) \cdot \nabla_z (X \circ \Phi^{-1}(Z; A)) \cdot \nabla_a \Phi^{-1}(Z; A). \end{aligned}$$

avec

$$\nabla_a \Phi^{-1}(Z; A) = \nabla_a \Phi(Z; -A) = [B_y(Z), -B_x(Z)]^T.$$

Il propose d'appliquer plusieurs cette correction, mais il ne précise pas le nombre de fois à mon avis cela est arbitraire et correspond à l'intégration dans le temps de la fonction de courant ?

Ainsi dans cet article, il propose une correction qui vérifie des contrainte physique (*divergence-free*). La régularisation sur le déplacement avec un prior doit être pris en compte et en pratique il préfère utiliser une régularisation de type Thikhonov sur les déformations.

## 5 Notes personnelles

Quelques remarques personnelles sur ces trois développements

- Ravela 2007 [8] : Les travaux de Ravela sont intéressants dans le sens où ils peuvent être combinés avec des méthodes classiques en assimilation de données. Néanmoins, la correction de position est basée sur un problème de minimisation non linéaire. Dans les tests que j'ai pu réaliser pour l'étape d'alignement, j'ai pu constater que la solution reste souvent dans des minima locaux. De plus, il y a une étape de calibration des hyperparamètres pour la régularisation. Si cela est mal réalisé, on se retrouve facilement avec des champs de déplacement qui ne sont pas réguliers. Dans notre cas de simulations sans maillage, on peut potentiellement utiliser la même méthode sur une grille régulière pour obtenir  $\mathbf{q}$  puis appliquer par interpolation la déformation. L'étape qui suit serait la même qu'auparavant, c'est-à-dire corriger les intensités. Cependant, on peut espérer avoir une discrétisation plus cohérente avec la solution analysée.
- Rosenthal 2017 [9] : Dans cet article, l'objectif est similaire à celui de Ravela, c'est-à-dire de modifier l'ébauche pour réduire l'erreur avec les données. Ici, la transformation appliquée doit être incompressible. Ainsi, ils vont chercher à déterminer les paramètres d'une fonction de courant qui est ensuite intégrée pour mettre à jour l'ébauche. Pour notre cas de méthode particulière, cet article présentant des mises à jour de l'ébauche est intéressant. L'intérêt ici par rapport à l'article de Ravela est que le transport se fait de manière physiquement consistante. L'application pour la méthode VIC que j'utilise peut être assez proche ici. On

---

1. Il y a une erreur (il me semble) dans leur article

appliquerait la fonction courant déterminée sur les particules de l'ébauche, mettant à jour l'ensemble d'ébauche. Puis, en travaillant sur l'ensemble transporté, on applique également les filtres d'ensemble classiques (Part-EnKF par exemple). Pour un cas plus général, par exemple en mécanique des solides avec la méthode MPM ou DEM, pourrait-on imaginer utiliser le même type de transformation ? En MPM, on peut imaginer appliquer une transformation, tout comme en VIC, les points matériau peuvent être facilement déplacés, à voir sur les conditions d'admissibilité cinématique. Pour la DEM, cela est plus complexe. En effet, les contraintes sur la mise à jour concernent la position des particules les unes par rapport aux autres. Ainsi, la mise à jour se fait par la détermination d'une force qui va être régularisée sur les interpénétrations et intégrée dans le schéma temporel. Pourrait-on imaginer réaliser une correction de cette force pour pouvoir modifier la trajectoire des particules sur un intervalle de temps fixe ?

- Bocquet 2023 [2] : Cet article propose une méthode qui applique le transport optimal à l'assimilation de données en introduisant deux champs supplémentaires pour traiter des cas où les masses de densité diffèrent. L'objectif est de travailler avec des observations partielles de champs et des observations bruitées, en corrigeant à la fois la position et les poids dans un espace de Riemann associé à la distance de Wasserstein. Ici, le choix est fait de travailler sur des discrétisations eulériennes pour définir les états  $\mathbf{x}^b, \mathbf{x}^o, \mathbf{x}^a$ . Dans l'article les discrétisation des états  $\mathbf{x}$  sont toutes les mêmes. Dans notre applications aux méthodes sans maillage, on connaît la discrétisation de  $\mathbf{x}^b$  uniquement. Cette article travaille sur des discrétisation de densité de champ, ainsi il me semble possible de pouvoir travailler directement sur les distributions *empiriques* de particules c'est dire  $P_i = \sum_p \Gamma_p \delta_{z_p}$ . Pour définir la discrétisation de  $\mathbf{x}^o$ , on pourrait choisir une discrétisation régulière arbitraire. En effet, il faudrait éviter de prendre uniquement le support des observations, le risque étant d'avoir un déséquilibre de masse. Finalement, l'étape de barycentre pourrait être obtenu par minimisation d'une fonction coût de type  $L(\mathbf{x}) = W(\mathbf{x}^o, \mathbf{x}) + W(\mathbf{x}^b, \mathbf{x})$ , en partant d'une discrétisation Lagrangienne du champ (par exemple en partant de  $\mathbf{x}^b$ ) et minimiser alternativement en fonction de  $z^p$  et  $\Gamma^p$ . L'objectif étant de suivre au mieux la géodésic. On pourrait s'inspirer des travaux de Cuturi et al. [3] ou bien de proposer une méthode de minimisation par gradient flow par exemple [6]. La formulation dynamique du problème de transport optimal pourrait être intéressante aussi car elle permet de déterminer un champ de vitesse, mais je ne crois pas encore avoir acquis l'ensemble des subtilités dans ce cas (voir chapitre 7 de [7]). Egalement je viens de trouver des formulations de MPM qui mêle MPM et OT [5].

## Références

- [1] Mathieu Blondel, Vivien Seguy, and Antoine Rolet. Smooth and Sparse Optimal Transport, February 2018. arXiv :1710.06276 [cs, stat].
- [2] Marc Bocquet, Pierre Vanderbecken, Alban Farchi, Joffrey Dumont Le Brazidec, and Yelva Roustan. *Bridging classical data assimilation and optimal transport*. December 2023.
- [3] Marco Cuturi and Arnaud Doucet. Fast Computation of Wasserstein Barycenters, June 2014. arXiv :1310.4375 [stat].
- [4] Nelson Feyeux. Transport optimal pour l'assimilation de données images. page 182.
- [5] B. Li, F. Habbal, and M. Ortiz. Optimal transportation meshfree approximation schemes for fluid and plastic flows. *International Journal for Numeri-*

- cal Methods in Engineering*, 83(12) :1541–1579, 2010. \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.2869>.
- [6] Antoine Liutkus, Umut Şimşekli, Szymon Majewski, Alain Durmus, and Fabian-Robert Stöter. Sliced-Wasserstein Flows : Nonparametric Generative Modeling via Optimal Transport and Diffusions, June 2019. arXiv :1806.08141 [cs, stat].
  - [7] Gabriel Peyré and Marco Cuturi. Computational Optimal Transport, March 2020. arXiv :1803.00567 [stat].
  - [8] Sai Ravela, Kerry Emanuel, and Dennis McLaughlin. Data assimilation by field alignment. *Physica D : Nonlinear Phenomena*, 230(1-2) :127–145, June 2007.
  - [9] W. Steven Rosenthal, Shankar Venkataramani, Arthur J. Mariano, and Juan M. Restrepo. Displacement data assimilation. *Journal of Computational Physics*, 330 :594–614, February 2017.
  - [10] Yue Ying. A Multiscale Alignment Method for Ensemble Filtering with Displacement Errors. *Monthly Weather Review*, 147(12) :4553–4565, December 2019. Publisher : American Meteorological Society Section : Monthly Weather Review.
  - [11] Yue Ying, Jeffrey L. Anderson, and Laurent Bertino. Improving Vortex Position Accuracy with a New Multiscale Alignment Ensemble Filter. *Monthly Weather Review*, 151(6) :1387–1405, June 2023.