Utilisation du transport optimal pour l'assimilation de données

Nelson Feyeux, Maëlle Nodet, Arthur Vidard

Utilisation du transport optimal pour l'assimilation de données

Nelson FEYEUX, Maëlle NODET, Arthur VIDARD

Assimilation de données océanographique

L'assimilation de données consiste à partir d'un modèle $\mathcal{M}$ d'évolution d'une variable $\mathbf{x}$, et d'observations **partielles** $(\mathbf{y}_i^o)$ de la variable $\mathbf{x}$, de retrouver la condition initiale $\mathbf{x}_0$. Le principe de l'assimilation de données variationnelle consiste à minimiser une fonction coût $\mathcal{J}$ définie par

$$\min_{\mathbf{x}_0} \mathcal{J}(\mathbf{x}_0) = \underbrace{\frac{1}{2}\sum_i d\Big(\mathcal{H}_i(\mathbf{x}_0), \mathbf{y}_i^o\Big)^2}_{\approx\, d\left(\mathbf{x}_0, \mathbf{x}_0^t\right)^2} + \frac{\gamma}{2} d\Big(\mathbf{x}_0, \mathbf{x}_0^b\Big)^2, \tag{1}$$

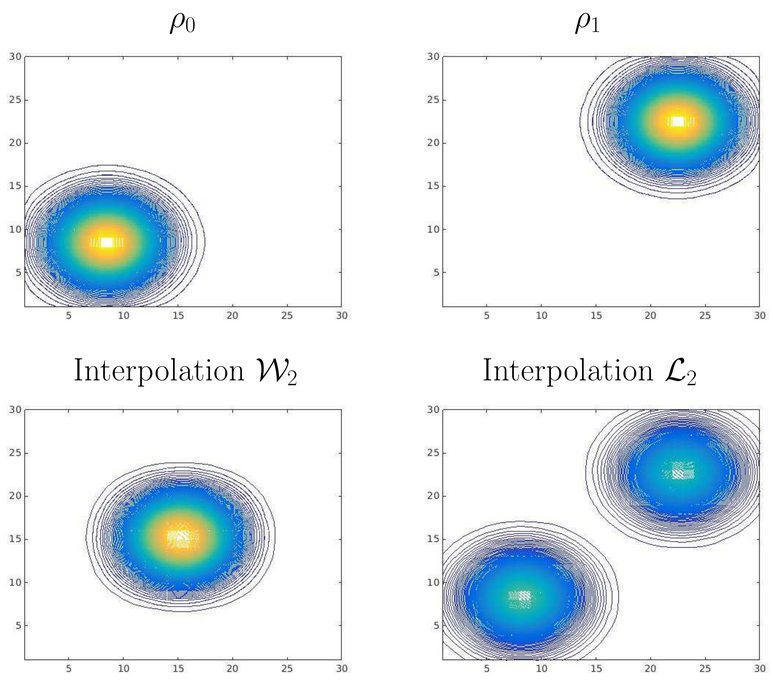
Généralement la distance d est une distance $\mathcal{L}_2$ pondérée. Or, en utilisant cette distance, l'assimilation de données variationnelle peine à faire face aux erreurs de déplacement. En prenant des distances tenant mieux en compte des spécificités des densités, on espère aboutir à une assimilation variationnelle rendant un état plus réaliste. C'est pourquoi on s'intéresse à la distance de Wasserstein $d = \mathcal{W}_2$.

Intérêt du transport optimal dans le cas d'erreur de déplacement.

Connaissant deux densités $\rho_0(x)$ et $\rho_1(x)$, la **distance de Wasserstein** $\mathcal{W}_2(\rho_0, \rho_1)$ est définie par

$$\mathcal{W}_2^2(\rho_0, \rho_1) := \inf_{\substack{(\rho(t,x), v(t,x)) \\ \partial_t \rho + \mathrm{div}(\rho v) = 0 \\ \rho(0,x) = \rho_0(x), \rho(1,x) = \rho_1(x)}} \int\int_{[0,1]\times\Omega} \rho(t,x)|v(t,x)|^2 \,\mathrm{d}t\mathrm{d}x.$$

Le $\rho(t,x)$ réalisant le minimum suit la **géodésique** dans l'espace de Wasserstein. Ci-dessous l'exemple d'interpolation pour la distance de Wasserstein et pour la distance $\mathcal{L}_2$ de ρ_0 et ρ_1 différant d'un déplacement.



Pour que la distance de Wasserstein soit définie, il faut que $\rho_0 \geq 0$, $\rho_1 \geq 0$ et $\int_\Omega \rho_0 = \int_\Omega \rho_1$.

Méthode du gradient avec la distance de Wasserstein

On se place dans le cas où $\mathbf{x}_0$, $\mathcal{H}_i(\mathbf{x}_0)$ et $\mathbf{y}_i^o$ sont des mesures de probabilité, pour tout i. Alors, la fonction coût écrite avec la distance de Wasserstein est

$$\mathcal{J}_W(\mathbf{x}_0) = \frac{1}{2}\sum_i \mathcal{W}_2^2\Big(\mathcal{H}_i(\mathbf{x}_0), \mathbf{y}_i^o\Big)$$

La méthode du gradient consiste à trouver $\mathrm{grad}\mathcal{J}_W(\mathbf{x}_0)$ et à utiliser l'algorithme

$$\mathbf{x}_0^{n+1} = \mathbf{x}_0^n - \alpha\, \mathrm{grad}\mathcal{J}_W(\mathbf{x}_0^n)$$

de sorte que $\mathcal{J}_W(\mathbf{x}_0^{n+1}) < \mathcal{J}_W(\mathbf{x}_0^n)$. Pour trouver le gradient, il faut calculer la **différentielle** et choisir un **produit scalaire**. En effet, $\mathrm{grad}\mathcal{J}(\mathbf{x}_0)$ est l'élément tel que pour tout η, $(\eta, \mathrm{grad}\mathcal{J}(\mathbf{x}_0)) = D\mathcal{J}[\mathbf{x}_0].\eta$.

La différentielle :

On différentie $\mathcal{J}_W$ en calculant $\mathcal{J}_W(\mathbf{x}_0 + \epsilon\eta)$:

- $\mathcal{H}_i(\mathbf{x}_0 + \epsilon\eta) = \mathcal{H}_i(\mathbf{x}_0) + \epsilon\, L[t_i, \mathbf{x}_0].\eta$ avec L le modèle tangent.
- $\frac{1}{2}\mathcal{W}_2^2(\mathbf{y} + \epsilon\mu, \mathbf{y}') = \epsilon\langle\mu, \psi\rangle_2$ avec ψ le **potentiel de Kantorovich** du transport optimal entre $\mathbf{y}$ et $\mathbf{y}'$.

On trouve ainsi que

$$\mathrm{D}\mathcal{J}_W[\mathbf{x}_0].\eta = \sum_i \Big\langle L[t_i, \mathbf{x}_0].\eta \ ,\ \psi_i \Big\rangle_2$$

avec ψ_i le potentiel de Kantorovich du transport optimal entre $\mathcal{H}_i(\mathbf{x}_0)$ et $\mathbf{y}_i^o$.

Le produit scalaire :

Utiliser le produit scalaire $\mathcal{L}_2$ pour définir le gradient n'est pas correct : la convergence de l'algorithme du gradient est très mauvaise. On utilise plutôt le produit scalaire $\mathcal{W}_2$, défini en $\mathbf{x}_0$ par :

$$(\eta, \eta') = \int_\Omega \mathbf{x}_0 \nabla\Phi \cdot \nabla\Phi',$$
$$\text{avec } \Phi,\ \Phi' \text{ tels que} \qquad \eta = -\mathrm{div}(\mathbf{x}_0\nabla\Phi), \qquad \eta' = -\mathrm{div}(\mathbf{x}_0\nabla\Phi')$$

Finalement, le gradient de $\mathcal{J}_W(\mathbf{x}_0)$ pour le produit scalaire $\mathcal{W}_2$ est, avec L^* le modèle *adjoint*,

$$\mathrm{grad}\mathcal{J}_W(\mathbf{x}_0) = -\mathrm{div}\left(\mathbf{x}_0 \nabla\left(\sum_i L^*[t_i, \mathbf{x}_0].\psi_i\right)\right).$$

Assimiler des variables n'étant pas des mesures de probabilité

Dans le cas où $\mathcal{H}_i(\mathbf{x}_0)$ et $\mathbf{y}_0$ sont des mesures de probabilité, mais *pas* $\mathbf{x}_0$, on ne peut pas écrire de terme d'ébauche. En effet, $\mathcal{W}_2^2(\mathbf{x}_0, \mathbf{x}_0^b)$ n'est pas défini. Par exemple soit le modèle d'évolution de Saint-Venant

$$\begin{aligned} &\partial_t h + \mathrm{div}(hu) = 0 \\ &\partial_t u + (u\cdot\nabla)u + g\nabla h = \mu\Delta u \end{aligned} \tag{2}$$

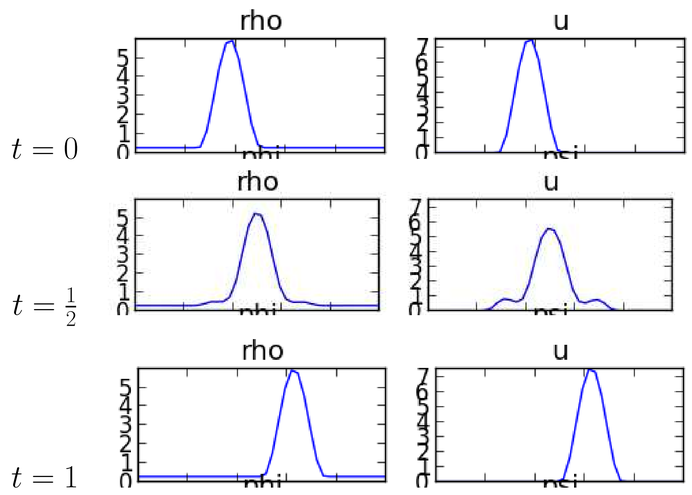
où l'on cherche à assimiler (h_0, u_0). La fonction coût s'écrit

$$\mathcal{J}(h_0, u_0) = \frac{1}{2}\sum_i \mathcal{W}_2^2\Big(\mathcal{H}_i(h_0, u_0), h_i^o\Big) + \gamma\mathcal{J}^b(h_0, u_0).$$

Qu'écrire pour le terme d'ébauche $\mathcal{J}^b(h_0, u_0)$? On ne peut pas écrire $\mathcal{W}_2^2((h_0, u_0), (h_0^b, u_0^b))$ puisque u_0 et u_0^b ne sont ni positifs ni de même intégrale. On ne peut pas ajouter un terme $\|u_0 - u_0^b\|_2^2$ puisque la convergence serait très mauvaise (mélange de distance de Wasserstein et de distance $\mathcal{L}_2$).
On utilise plutôt le terme d'ébauche suivant

$$J^b(h_0, u_0) = \inf_{\substack{\partial_t\rho + \mathrm{div}(\rho v) = 0 \\ \partial_t u + \mathrm{div}(\rho w) = 0 \\ \rho(0,x) = h_0, \rho(1,x) = h_0^b \\ u(0,x) = u_0, u(1,x) = u_0^b}} \int\int_{[0,1]\times\Omega} (|v|^2 + |w|^2)\,\rho \,\mathrm{d}t\mathrm{d}x, \tag{3}$$

Dans ce cas l'interpolation entre deux (h_0, u_0) et (h_1, u_1) décalés redonne bien un (h, u) entre les deux.



Le produit scalaire qu'on choisira sera

$$\left(\begin{pmatrix}\eta\\ v\end{pmatrix}, \begin{pmatrix}\eta'\\ v'\end{pmatrix}\right) = \int_\Omega h_0\nabla\Phi\cdot\nabla\Phi' + \int_\Omega h_0\nabla\Psi\cdot\nabla\Psi'$$

avec Φ, Φ', Ψ, Ψ' tels que

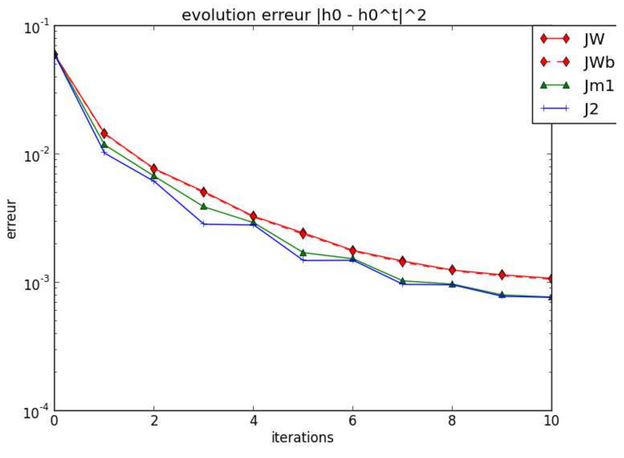
$$\begin{aligned} \eta &= -\mathrm{div}(h_0\nabla\Phi), & \eta' &= -\mathrm{div}(h_0\nabla\Phi') \\ v &= -\mathrm{div}(h_0\nabla\Psi), & v' &= -\mathrm{div}(h_0\nabla\Psi'). \end{aligned}$$

Difficultés liées au transport optimal

- La distance de Wasserstein n'est définie que pour des mesures de probabilité.
- Lorsque $\rho_0, \rho_1 \approx$ cte, l'interpolation $\mathcal{W}_2$ ressemble à l'interpolation $\mathcal{L}_2$...
- Lorsque $\mathcal{J}(h_0^n) \to \min_{h_0}\mathcal{J}(h_0)$, on a seulement $h_0^n \rightharpoonup \arg\min_{h_0}\mathcal{J}(h_0)$.
- Le temps de calcul numérique de la distance de Wasserstein est plus long que celui de la distance $\mathcal{L}_2$ [Peyré, Papadakis, Oudet, 2013].

Résultats et perspectives

Pour de premiers tests sur l'assimilation (2) on compare l'erreur $\|h_0 - h_0^t\|_2^2$ en utilisant $d = \mathcal{L}_2$ et $d = \mathcal{W}_2$. Les comportements sont corrects.



Il reste à implémenter l'ébauche (3).

Ce travail est supporté par la région Rhône-Alpes.