Optimal Transportation for Data Assimilation

Nelson Feyeux, Maëlle Nodet, Arthur Vidard

To cite this version:


HAL Id: hal-01349637
https://hal.archives-ouvertes.fr/hal-01349637
Submitted on 28 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
DATA ASSIMILATION

Let the model $\mathcal{M}$, for the atmosphere, the state $x(t,x,y)$ gathers the different variables:
- humidity $h(t,x,y)$;
- temperatures $T(t,x,y)$;
- pressure $p(t,x,y)$.

Can we estimate the initial condition $x_0$ of the system?

The Wasserstein cost function writes

$$W(q,x) = \inf_{\pi} \int d(x, y) \sqrt{\pi(x) \pi(y)}$$

with the Wasserstein distance $d(x,y)$.

Computing the Wasserstein distance is expensive [Peyré, Papadakis, Oudet, 2013];

the distance $d(x,y)$ is common for the distance $d$.

Relaxations of the latter constraint are possible, however complex;

when $\mathcal{J}(\rho) \to \min_{\rho} \mathcal{J}(\rho)$, then there is only weak convergence of $\rho_{\text{opt}}$ to $\rho_0$. oscillations or diverges can occur!

Computing the Wasserstein distance is expensive [Peyré, Papadakis, Oudet, 2013];

RESULTS ON A SHALLOW-WATER EQUATION

Let the model $\mathcal{M}$ be a Shallow-Water equation, with initial condition $(h_0, u_0)$.

$$\mathcal{M} : \begin{cases} \frac{\partial h}{\partial t} + \text{div}(hu) = 0 \\ \frac{\partial u}{\partial t} + u \cdot \nabla u = -g \nabla h. \end{cases}$$

We control the initial condition $h_0$ only, thanks to the Wasserstein cost function $\mathcal{J}_0$. We set $u_0 = 0$.

The observations of $h^\text{true}$ at times $t_i$.

Variational data assimilation consists in retrieving $x_0$ by minimizing

$$J(x_0) := \sum_i d^2(x_0, y_i) + \omega d(x_0, x_i)^2.$$ (1)

It is common for the distance $d$ to be a weighted $L^2$ distance. Our main goal to use the Wasserstein distance $W_2$ instead, which seems very interesting when dealing with dense data (see right panel). The Wasserstein cost function writes

$$J_0(x_0) := \sum_i W_2^2(x_0, y_i) + \omega W_2^2(x_0, x_i)^2.$$ (2)

SPECIFICITIES ON USING THE WASSERSTEIN DISTANCE

- The Wasserstein distance is only defined for probability measures, i.e. $\rho \geq 0$.
- $\mathcal{J}_0$ interpolation works well if $\rho_0$ and $\rho_1$ are of distinct support;
- when $\mathcal{J}(\rho) \to \min_{\rho} \mathcal{J}(\rho)$, then there is only weak convergence of $\rho_{\text{opt}}$ to $\rho_0$.
- Computing the Wasserstein distance is expensive [Peyré, Papadakis, Oudet, 2013].

OPTIMAL TRANSPORTATION AND THE WASSERSTEIN DISTANCE

For two functions $\rho_1(x)$ and $\rho_2(x)$, the square of the Wasserstein distance $W_2(\rho_1, \rho_2)$ is defined as the minimal kinetic energy necessary to transport $\rho_1$ to $\rho_2$,

$$W_2(\rho_1, \rho_2)^2 = \frac{1}{2} \int h_t \int d\rho^2 \sqrt{\rho_1(x) \rho_2(y)}.$$ (3)

For the Wasserstein distance to be well-defined, one needs $\rho_1 \geq 0, \rho_2 \geq 0$ and $\int \rho_1 = \int \rho_2 = 1$.

Average w.r.t the Wasserstein distance

The average, or barycenter, minimizes $W_2(\rho, \rho_0)^2$ and $W_2(\rho, \rho_1)^2$. It is also the optimal $\rho$ in the definition of $W_2(\rho_1, \rho_2)^2$ t time $t = 1/2$.

EXAMPLE OF USE OF THE WASSERSTEIN distance

The model $\mathcal{M}$ is a Shallow-Water model, with initial condition $(h_0, u_0)$.

$\mathcal{M} : \begin{cases} \frac{\partial h}{\partial t} + \text{div}(hu) = 0 \\ \frac{\partial u}{\partial t} + u \cdot \nabla u = -g \nabla h. \end{cases}$

We control the initial condition $h_0$ only, thanks to the Wasserstein cost function $\mathcal{J}_0$. We set $u_0 = 0$.

The observations of $h^\text{true}$ at times $t_i$.

True and background initial conditions.

- The minimization of $\mathcal{J}_0$ is performed through a gradient descent, using the Wasserstein gradient, arising from the use of the following Wasserstein scalar product depending on $\rho_0$.

$$\mathcal{J}(\rho) = \frac{1}{2} \int h_t \int d\rho^2 \sqrt{\rho_1(x) \rho_2(y)}.$$ (3)

For $\eta, \eta'$ s.t. $\int \eta = 1$

Let $\Phi, \Phi'$ s.t. $-\text{div} (\rho_0 \nabla \Phi) = \eta$ (with Neumann BC)

$$\text{div} (\rho_0 \nabla \Phi') = \eta'$$

Then $\langle \eta, \eta' \rangle_{\mathcal{J}} = \int h_t \int \rho_0 \nabla \Phi \cdot \nabla \Phi'.d\rho.$