Optimal Transportation for Data Assimilation
Nelson Feyeux, Maëlle Nodet, Arthur Vidard

To cite this version:
Nelson Feyeux, Maëlle Nodet, Arthur Vidard. Optimal Transportation for Data Assimilation. 5th International Symposium for Data Assimilation (ISDA 2016), Jul 2016, Reading, United Kingdom., 2016. hal-01349637

HAL Id: hal-01349637
https://hal.science/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.
Optimal Transportation for Data Assimilation

Nelson FEYEUX, Maëlle NODET, Arthur VIDARD
Inria & Université Grenoble-Alpes

DATA ASSIMILATION

Given
- a physical system and its state \( x(t, x, y) \);
- partial observations of the system \( y_i(t) \);
- a (numeric) model \( M \) simulating the evolution of \( x \);

Can we estimate the initial condition \( x_0 \) of the system?

\[ J(x_0) = \sum_i d\left( H_i(M(x_0)), y_i^0 \right)^2 + \omega d\left( x_0, x^i_0 \right)^2. \]  

(1)

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

\[ J_W(x_0) = \sum_i W_i(H_i(M(x_0)), y_i^0)^2 + \omega W_i(x_0, x_i^0)^2. \]  

(2)

RESULTS ON A SHALLOW-WATER EQUATION

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

\[ \begin{align*}
\frac{\partial h}{\partial t} + \text{div}(hu) &= 0 \\
\frac{\partial u}{\partial t} + u \cdot \text{grad}u &= -g \text{V}h.
\end{align*} \]

We control the initial condition \( h_0 \) only, thanks to the Wasserstein cost function \( J_W \). We set \( u_0 = 0 \).

\[ t_0 \quad t_1 \quad t_2 \quad t_3 \quad t_4 \quad t_5 \]

The observations of \( h^{\text{true}} \) at times \( t_i \)

\[ \begin{align*}
\text{background} & \quad \text{true} \\
\end{align*} \]

True and background initial conditions

SPECIFICITIES ON USING THE WASSERSTEIN DISTANCE

- The Wasserstein distance is only defined for probability measures, i.e. \( \rho \) s.t.
  \[ \rho \geq 0 \quad \text{and} \quad \int \rho = 1. \]

  Relaxations of the latter constraint are possible, however complex;
  - the \( W \) interpolation works well if \( \rho_0 \) and \( \rho_1 \) are of direct support;
  - when \( J(\rho^0) \rightarrow \min_{\rho\geq0} J(\rho) \), then there is only a weak convergence of \( \rho^0 \) to \( \rho^\text{opt} \): oscillations or diracs can occur!
  - Computing the Wasserstein distance is expensive [Peyré, Papadakis, Oudet, 2013];

- The minimization of \( J_W \) is performed through a gradient descent, using the Wasserstein gradient, arising from the use of the following Wasserstein scalar product depending on \( \rho_0 \):
  \[ \begin{align*}
  \left( \rho_0, \eta \right)_1 &= \int_\Omega \rho_0 \eta \text{div} \Phi \text{d}x \\
  \left( \rho_0, \eta \right)_2 &= \int_\Omega \rho_0 \eta \text{div} \Phi \text{d}x + \int_\Omega \rho_0 \eta \Phi \text{d}x \quad \text{(with Neumann BC)}
  \end{align*} \]

  \[ \begin{align*}
  \text{For} \; \eta, \eta' \; \text{s.t.} \; \int \eta = \int \eta' = 0 \\
  \text{Let} \; \Phi, \Phi' \; \text{s.t.} \; -\text{div}(\rho_0 \Phi) = \eta \; \text{(with Neumann BC)} \\
  \text{and} \; -\text{div}(\rho_0 \Phi') = \eta',
  \end{align*} \]

  \[ \begin{align*}
  \text{Then} \; \langle \eta, \eta' \rangle_1 &= \int_\Omega \rho_0 \eta \text{div} \Phi \cdot \eta' \text{d}x \\
  \text{and} \; \langle \eta, \eta' \rangle_2 &= \int_\Omega \rho_0 \eta \text{div} \Phi \cdot \eta' \text{d}x + \int_\Omega \rho_0 \eta \Phi \cdot \eta' \text{d}x.
  \end{align*} \]

OPTIMAL TRANSPORTATION AND THE WASSERSTEIN DISTANCE

For two functions \( \rho_0(x) \) and \( \rho_1(x) \), the square of the Wasserstein distance \( W_2(\rho_0, \rho_1) \) is defined as the minimal kinetic energy necessary to transport \( \rho_0 \) to \( \rho_1 \):

\[ W_2(\rho_0, \rho_1)^2 = \inf_{\|\gamma\|_2 = 1} \frac{1}{2} \int_\Omega \rho(x)^2 \text{d}x. \]

\[ \rho(0, x) = \rho_0(x), \rho(t, x) = \rho_1(x) \]

For the Wasserstein distance to be well-defined, one needs

\[ \rho_0 \geq 0, \rho_1 \geq 0 \quad \text{and} \quad \int \rho_0 = \int \rho_1 = 1. \]

Average w.r.t the Wasserstein distance

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

\[ W_2 \] average

Example of use of the Wasserstein distance

(Source : Cuturi, Doucet, 2013)

Results :

Analysis \( h_0 \) of the assimilation when using the Euclidean \( (L^2) \) or the Wasserstein \( (W_2) \) distance.

Values of \( h \) and \( u \) for the background and true states, as well as analysis for Euclidean and Wasserstein distances, at time \( t = t_0 \)

References

Nelson Feyieux, Maëlle Nodet, Arthur Vidard. Optimal Transport for Data Assimilation. 2016. <hal-01342193>

This work has been supported by the region Rhône-Alpes.