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

To cite this version:
Nelson Feyieux, 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.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**

Given:
- a physical system and its state \( x(t, x, y) \);
- partial observations of the system \( \gamma(t) \);
- a (numerical) model \( M \) simulating the evolution of \( x \);
- \( F \), for the atmosphere, the state \( x(t, x, y) \) gathers the different variables:
  - humidity \( H(t, x, y) \);
  - velocities \( u(t, x, y) \);
  - temperature \( T(t, x, y) \);
  - pressure \( p(t, x, y) \).

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

Variational data assimilation consists in retrieving \( x_0 \) by minimizing

\[
\mathcal{J}(x_0) := \sum_i d\left( H(M(x_0)) \gamma_i \right)^2 + \omega d(x_0, x_i)^2.
\]

(1)

It is common for the distance \( d \) to be a weighted \( L^2 \) distance. Our main goal is 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

\[
\mathcal{J}(x_0) := \sum_i W_2(H(M(x_0)), \gamma_i)^2 + \omega W_2(x_0, x_i)^2.
\]

(2)

**RESULTS ON A SHALLOW-WATER EQUATION**

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

\[
M : \begin{cases}
\partial_t h + \text{div}(hu) = 0 \\
\partial_t u + u \cdot \nabla u = -g \nabla h.
\end{cases}
\]

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

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

**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 \( W_2 \) interpolation works well if \( \rho_0 \) and \( \rho_1 \) are of distinct support;
  - when \( \mathcal{J}(\rho) \rightarrow \min \mathcal{J}(\rho_0) \), then there is only weak convergence of \( \rho_t \) to \( \rho_0^\text{opt} \): oscillations or diracs can occur!
  - Computing the Wasserstein distance is expensive [Peyré, Papadakis, Oudet, 2013].

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

  \[
  (\rho_1, \rho_2) = \frac{1}{2} \int \rho_1 \rho_2 d|x|.
  \]

  For \( \eta, \eta' \) s.t.
  \[
  \eta = \int_{\Omega} \eta' = 0.
  \]

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

  \[
  -\text{div}(\rho_0 \nabla \Phi') = \eta'.
  \]

  Then
  \[
  \eta \Phi'|\Omega = \int_{\Omega} \rho_0 \nabla \Phi \cdot \nabla \Phi' d|x|.
  \]

**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 \( \rho_1 \),

\[
W_2(\rho_0, \rho_1)^2 := \inf \left( \frac{1}{2} \int |\rho'(x)|^2 d|x| \right)
\]

\[
\rho(0, x) = \rho_0(x), \rho(1, x) = \rho_1(x)
\]

For the Wasserstein distance to be well-defined, one needs \( \rho_0 \geq 0, \rho_1 \geq 0 \) and \( \int \rho_0 = \int \rho_1 = 1 \).

Average w.r.t the Wasserstein distance

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

Example of use of the Wasserstein distance

(Source: Urban, Doucet, Fast computation of Wasserstein barycenters)

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 \( t = t_0 \).

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