

Optimal transport for data assimilation from images

Nelson FEYEUX

INRIA LJK Grenoble - Maelle NODET - Arthur VIDARD

16 Octobre 2015 - Workshop Optimal Transport 2015



LABORATOIRE
JEAN KUNTZMANN
MATHÉMATIQUES APPLIQUÉES • INFORMATIQUE



- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

Definition of the Wasserstein distance

- Ω is a closed bounded set of \mathbb{R}^d ,

Definition of the Wasserstein distance

- Ω is a closed bounded set of \mathbb{R}^d ,
- $\mathcal{P}(\Omega) = \{\rho: \rho \geq 0, \int_{\Omega} \rho = 1\}$,

Definition of the Wasserstein distance

- Ω is a closed bounded set of \mathbb{R}^d ,
- $\mathcal{P}(\Omega) = \{\rho: \rho \geq 0, \int_{\Omega} \rho = 1\}$,
- Let $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$ and let's draw a continuous path $\rho(t)$ between them.



Definition of the Wasserstein distance

- Ω is a closed bounded set of \mathbb{R}^d ,
- $\mathcal{P}(\Omega) = \{\rho: \rho \geq 0, \int_{\Omega} \rho = 1\}$,
- Let $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$ and let's draw a continuous path $\rho(t)$ between them.



There exists a velocity field $v(t, x)$ associated to $\rho(t)$ s.t.

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0 \quad \text{and} \quad v \cdot \vec{n} = 0 \quad \text{on} \quad \partial \Omega$$

Definition of the Wasserstein distance

- Ω is a closed bounded set of \mathbb{R}^d ,
- $\mathcal{P}(\Omega) = \{\rho: \rho \geq 0, \int_{\Omega} \rho = 1\}$,
- Let $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$ and let's draw a continuous path $\rho(t)$ between them.



There exists a velocity field $v(t, x)$ associated to $\rho(t)$ s.t.

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0 \quad \text{and} \quad v \cdot \vec{n} = 0 \quad \text{on } \partial\Omega$$

Proof : at each time t , find Φ s.t.

$$\begin{cases} -\operatorname{div}(\rho \nabla \Phi) = \frac{\partial \rho}{\partial t} \\ \frac{\partial \Phi}{\partial \vec{n}} = 0 \quad \text{on } \partial\Omega. \end{cases} \quad \Phi = -\Delta_{\rho}^{-1} \left(\frac{\partial \rho}{\partial t} \right)$$

Definition of the Wasserstein distance

Definition of the Wasserstein distance

- Let $\rho_0, \rho_1 \in P(\Omega)$, the Wasserstein distance is

$$\mathcal{W}_2^2(\rho_0, \rho_1) := \inf_{(\rho, v) \in C(\rho_0, \rho_1)} \iint_{[0,1] \times \Omega} \rho(t, x) |v(t, x)|^2 dt dx$$

- with

$$C(\rho_0, \rho_1) := \left\{ (\rho, v) \text{ s.t. } \begin{array}{l} \rho(0) = \rho_0, \rho(1) = \rho_1, \\ \partial_t \rho + \operatorname{div}(\rho v) = 0, \\ v \cdot \vec{n} = 0 \text{ on } \partial\Omega \end{array} \right\}$$

Video example !

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \inf_{\partial_t \rho + \operatorname{div}(\rho v) = 0} \iint_{[0,1] \times \Omega} \rho |v|^2 dt dx.$$

Video example !

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \inf_{\partial_t \rho + \operatorname{div}(\rho v) = 0} \iint_{[0,1] \times \Omega} \rho |v|^2 dt dx.$$

Minimizer properties

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \inf_{\partial_t \rho + \operatorname{div}(\rho v) = 0} \iint_{[0,1] \times \Omega} \rho |v|^2 dt dx.$$

Let (ρ, v) be a minimizer :

- $v = \nabla \Phi$

$$v = \nabla \Phi + w \text{ with } \operatorname{div}(\rho w) = 0 \Rightarrow w = 0.$$

Minimizer properties

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \inf_{\partial_t \rho + \operatorname{div}(\rho v) = 0} \iint_{[0,1] \times \Omega} \rho |v|^2 dt dx.$$

Let (ρ, v) be a minimizer :

- $v = \nabla \Phi$

$$v = \nabla \Phi + w \text{ with } \operatorname{div}(\rho w) = 0 \Rightarrow w = 0.$$

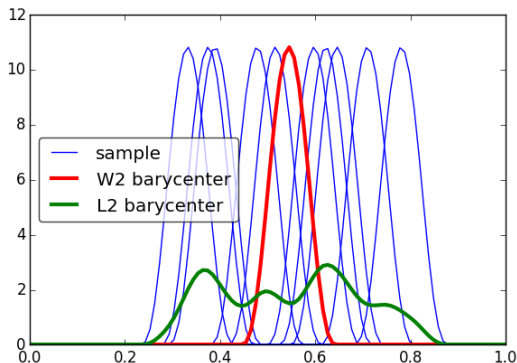
- The equations for the minimizers are

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho v) = 0 \\ \partial_t v + v \cdot \nabla v = 0 \end{cases} \quad \begin{cases} \partial_t \rho + \operatorname{div}(\rho \nabla \Phi) = 0 \\ \partial_t \Phi + \frac{|\nabla \Phi|^2}{2} = 0 \end{cases}$$

- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

Wasserstein distance has a good behavior under displacement errors.

$$\inf_{\rho} \sum_i \mathcal{W}_2^2(\rho, \rho^i) \quad \text{vs} \quad \inf_{\rho} \sum_i \mathcal{L}_2^2(\rho, \rho^i) = \|\rho - \rho^i\|_2^2$$



- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

Weather forecasting



$$\mathbf{X} = \begin{pmatrix} \text{Temperature } T \\ \text{Velocity } v \\ \text{Pressure } p \\ \text{Density } \rho \\ \dots \end{pmatrix}$$

The state $\mathbf{X}(t, x)$ is defined everywhere in the atmosphere/ocean!

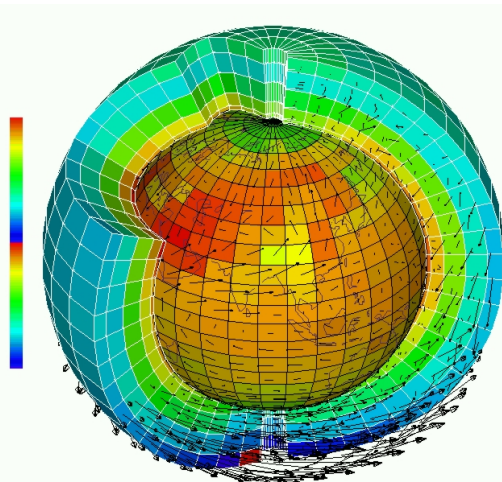
A model... and some computers !

A model...

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0 \\ \frac{DT}{Dt} = K_T \Delta T \\ \frac{Du}{Dt} = 2\vec{\Omega} \wedge u - \frac{1}{\rho} \nabla p + \nu \Delta u \\ \frac{\partial p}{\partial z} = -\rho g \end{cases}$$

$$\mathbf{X}(t) = \mathcal{M}(\mathbf{X}_0, t)$$

Numerical simulations



Observation of the ocean...

How to recover \mathbf{X}_0 from this ?

Retrouver la condition initiale

Goal : find \mathbf{X}_0 with the help of the observations \mathbf{Y}^o .

$$\min_{\mathbf{X}_0} \mathcal{J}(\mathbf{X}_0) = d\left(\mathcal{HM}(\mathbf{X}_0), \mathbf{Y}^o\right)$$

Actually, $d = \mathcal{L}^2$ (weighted). Let's try with $d = \mathcal{W}_2$!

Retrouver la condition initiale

Goal : find \mathbf{X}_0 with the help of the observations \mathbf{Y}^o .

$$\min_{\mathbf{X}_0} \mathcal{J}(\mathbf{X}_0) = d\left(\mathcal{HM}(\mathbf{X}_0), \mathbf{Y}^o\right) + d(\mathbf{X}_0, \mathbf{X}_0^b)$$

Actually, $d = \mathcal{L}^2$ (weighted). Let's try with $d = \mathcal{W}_2$!

- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

The Wasserstein cost

$$\mathcal{J}_{\mathcal{W}}(\mathbf{X}_0) = \underbrace{\sum_i \mathcal{W}_2^2(\mathcal{H}_i \mathcal{M} \mathbf{X}_0, \mathbf{Y}_i^{\text{obs}})}_{\mathcal{J}_o, \text{ distance to observations}} + \omega_b \underbrace{\mathcal{W}_2^2(\mathbf{X}_0, \mathbf{X}_0^b)}_{\mathcal{J}_b, \text{ distance to background}}$$

Problem : \mathbf{X}_0 , $\mathcal{H}_i \mathcal{M} \mathbf{X}_0$, $\mathbf{Y}_i^{\text{obs}}$ and \mathbf{X}_0^b must be in $P(\Omega)$.

The Wasserstein cost

$$\mathcal{J}_{\mathcal{W}}(\mathbf{X}_0) = \underbrace{\sum_i \mathcal{W}_2^2(\mathcal{H}_i \mathcal{M} \mathbf{X}_0, \mathbf{Y}_i^{\text{obs}})}_{\mathcal{J}_o, \text{ distance to observations}} + \omega_b \underbrace{\mathcal{W}_2^2(\mathbf{X}_0, \mathbf{X}_0^b)}_{\mathcal{J}_b, \text{ distance to background}}$$

Problem : \mathbf{X}_0 , $\mathcal{H}_i \mathcal{M} \mathbf{X}_0$, $\mathbf{Y}_i^{\text{obs}}$ and \mathbf{X}_0^b must be in $P(\Omega)$.

Solution :

- $\mathbf{X} \in P(\Omega)$ can represent water level or cloud distribution.
- \mathcal{H}_i is identity (observation on \mathbf{X} is total),
- The model \mathcal{M} is s.t. $\mathbf{X} \in P(\Omega) \Rightarrow \mathcal{M} \mathbf{X} \in P(\Omega)$.

First example

$$\mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \mathcal{W}_2^2(\mathcal{M}_i \rho_0, \rho_i^{\text{obs}}) + \omega_b \mathcal{W}_2^2(\rho_0, \rho_0^b)$$

with model $\mathcal{M} : \partial_t \rho + \text{div}(\rho u) = 0$ with a known $u(t, x)$.

- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

Computation of the Wasserstein distance

- In 1D there is a formula for computing $\mathcal{W}_2^2(\rho_0, \rho_1)$ (yet, it needs inversions thus somehow approximations).

Computation of the Wasserstein distance

- In 1D there is a formula for computing $\mathcal{W}_2^2(\rho_0, \rho_1)$ (yet, it needs inversions thus somehow approximations).
- Primal-dual algorithm, which is a good approximation of $\mathcal{W}_2^2(\rho_0, \rho_1)$.
 - Only an approximation (around the minimum of $\mathcal{J}_{\mathcal{W}}$, weird things can happen)
 - Stopping criterium to control the error ???

- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

Minimization

$$\text{Minimize } \mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \mathcal{W}_2^2(\mathcal{M}_i \rho_0, \rho_i^{\text{obs}}) + \omega_b \mathcal{W}_2^2(\rho_0, \rho_0^b)$$

Solve numerically by doing several iterations

$$\rho_0^{n+1} = \rho_0^n + \alpha^n d^n$$

- $d^n = -\text{grad} \mathcal{J}_{\mathcal{W}}(\rho_0^n)$ (gradient descent)
- or a function of $\text{grad} \mathcal{J}_{\mathcal{W}}(\rho_0^n)$ (quasi-Newton, BFGS, ...)

Definition of gradient

$$\forall \eta \in TP_{\rho_0}, \quad \mathcal{J}_{\mathcal{W}}(\rho_0 + \epsilon \eta) = \mathcal{J}_{\mathcal{W}}(\rho_0) + \epsilon \langle \text{grad} \mathcal{J}_{\mathcal{W}}, \eta \rangle + o(\epsilon)$$

- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

Wasserstein inner product

The \mathcal{L}_2 inner product between $\eta, \eta' \in TP_\rho$ is

$$\langle \eta, \eta' \rangle_2 = \int_{\Omega} \eta \eta' \, dx$$

Wasserstein inner product

The \mathcal{L}_2 inner product between $\eta, \eta' \in TP_\rho$ is

$$\langle \eta, \eta' \rangle_2 = \int_{\Omega} \eta \eta' \, dx$$

The \mathcal{W}_2 inner product is :

$$\Phi := -\Delta_\rho^{-1} \eta, \quad \Phi' := -\Delta_\rho^{-1} \eta',$$

$$\langle \eta, \eta' \rangle_\rho = \int_{\Omega} \rho \nabla \Phi \cdot \nabla \Phi' \, dx.$$

Wasserstein inner product

The \mathcal{L}_2 inner product between $\eta, \eta' \in TP_\rho$ is

$$\langle \eta, \eta' \rangle_2 = \int_{\Omega} \eta \eta' \, dx$$

The \mathcal{W}_2 inner product is :

$$\Phi := -\Delta_\rho^{-1} \eta, \quad \Phi' := -\Delta_\rho^{-1} \eta',$$

$$\langle \eta, \eta' \rangle_\rho = \int_{\Omega} \rho \nabla \Phi \cdot \nabla \Phi' \, dx.$$

Note : this inner product is natural to the Wasserstein distance :

$$\mathcal{W}_2(\rho_0, \rho_1) = \inf_{\text{path } \rho(t) \text{ between } \rho_0 \text{ and } \rho_1} \int_0^1 \|\partial_t \rho\| \, dt$$

Gradient of $\mathcal{J}_{\mathcal{W}}$

Definition of gradient

$$\mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \mathcal{W}_2^2(\mathcal{M}_i \rho_0, \rho_i^{\text{obs}}) + \omega_b \mathcal{W}_2^2(\rho_0, \rho_0^b)$$

Gradient of $\mathcal{J}_{\mathcal{W}}$

Definition of gradient

$$\mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \mathcal{W}_2^2(\mathcal{M}_i \rho_0, \rho_i^{\text{obs}}) + \omega_b \mathcal{W}_2^2(\rho_0, \rho_0^b)$$

$$\mathcal{J}_{\mathcal{W}}(\rho_0 + \epsilon \eta) = \sum_i \mathcal{W}_2^2(\mathcal{M}_i(\rho_0 + \epsilon \eta), \rho_i^{\text{obs}}) + \omega_b \mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_0^b)$$

Gradient of $\mathcal{J}_{\mathcal{W}}$

Definition of gradient

$$\mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \mathcal{W}_2^2(\mathcal{M}_i \rho_0, \rho_i^{\text{obs}}) + \omega_b \mathcal{W}_2^2(\rho_0, \rho_0^b)$$

$$\mathcal{J}_{\mathcal{W}}(\rho_0 + \epsilon \eta) = \sum_i \mathcal{W}_2^2\left(\underbrace{\mathcal{M}_i(\rho_0 + \epsilon \eta)}_{\mathcal{M}_i \rho_0 + \epsilon \frac{\partial \mathcal{M}_i}{\partial \rho_0}(\rho_0) \cdot \eta + o(\epsilon)}, \rho_i^{\text{obs}}\right) + \omega_b \mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_0^b)$$

Gradient of $\mathcal{J}_{\mathcal{W}}$

Definition of gradient

$$\mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \mathcal{W}_2^2(\mathcal{M}_i \rho_0, \rho_i^{\text{obs}}) + \omega_b \mathcal{W}_2^2(\rho_0, \rho_0^b)$$

$$\mathcal{J}_{\mathcal{W}}(\rho_0 + \epsilon \eta) = \sum_i \mathcal{W}_2^2\left(\underbrace{\mathcal{M}_i(\rho_0 + \epsilon \eta)}_{\mathcal{M}_i \rho_0 + \epsilon \frac{\partial \mathcal{M}_i}{\partial \rho_0}(\rho_0) \cdot \eta + o(\epsilon)}, \rho_i^{\text{obs}}\right) + \omega_b \mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_0^b)$$

$$= \sum_i \mathcal{W}_2^2\left(\mathcal{M}_i \rho_0 + \epsilon \frac{\partial \mathcal{M}_i}{\partial \rho_0}(\rho_0) \cdot \eta, \rho_i^{\text{obs}}\right) + \omega_b \mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_0^b)$$

Value of $\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_1)$

Differentiation of Wasserstein distance

$$\frac{\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_1) - \mathcal{W}_2^2(\rho_0, \rho_1)}{2\epsilon} \rightarrow \int_{\Omega} \eta(x)\Phi(x) dx.$$

with the potential Φ being (equivalently) :

Value of $\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_1)$

Differentiation of Wasserstein distance

$$\frac{\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_1) - \mathcal{W}_2^2(\rho_0, \rho_1)}{2\epsilon} \rightarrow \int_{\Omega} \eta(x)\Phi(x) dx.$$

with the potential Φ being (equivalently) :

- The Kantorovich potential of the transport of ρ_0 to ρ_1 .

Value of $\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_1)$

Differentiation of Wasserstein distance

$$\frac{\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_1) - \mathcal{W}_2^2(\rho_0, \rho_1)}{2\epsilon} \rightarrow \int_{\Omega} \eta(x)\Phi(x) dx.$$

with the potential Φ being (equivalently) :

- The Kantorovich potential of the transport of ρ_0 to ρ_1 .
- If $T = \nabla\Psi$, then $\Phi = |x|^2/2 - \Psi$.

Value of $\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_1)$

Differentiation of Wasserstein distance

$$\frac{\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_1) - \mathcal{W}_2^2(\rho_0, \rho_1)}{2\epsilon} \rightarrow \int_{\Omega} \eta(x)\Phi(x) dx.$$

with the potential Φ being (equivalently) :

- The Kantorovich potential of the transport of ρ_0 to ρ_1 .
- If $T = \nabla\Psi$, then $\Phi = |x|^2/2 - \Psi$.
- If $(\rho(t, x), v(t, x))$ draws the geodesic between ρ_0 and ρ_1 ,
 $v(t, x) = -\nabla\Phi(t, x)$, $\Phi(x) := \Phi(0, x)$.

Proof

Kantorovich potential :

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \max_{\psi, \psi(x) + \psi^*(y) \leq |x-y|^2} \int \psi(x) \rho_0(x) dx + \int \psi^*(y) \rho_1(y) dy$$

Proof

Kantorovich potential :

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \max_{\Psi, \Psi(x) + \Psi^*(y) \leq |x-y|^2} \int \Psi(x) \rho_0(x) dx + \int \Psi^*(y) \rho_1(y) dy$$

$$\mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_1) \geq \int \Psi(x) (\rho_0 + \epsilon \eta) dx + \int \Psi^*(y) \rho_1(y) dy$$

Proof

Kantorovich potential :

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \max_{\Psi, \Psi(x) + \Psi^*(y) \leq |x-y|^2} \int \Psi(x) \rho_0(x) dx + \int \Psi^*(y) \rho_1(y) dy$$

$$\mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_1) \geq \int \Psi(x) (\rho_0 + \epsilon \eta) dx + \int \Psi^*(y) \rho_1(y) dy$$

$$\frac{\mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_1) - \mathcal{W}_2^2(\rho_0, \rho_1)}{\epsilon} \geq \int_{\Omega} \Psi \eta$$

Proof

Kantorovich potential :

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \max_{\Psi, \Psi(x) + \Psi^*(y) \leq |x-y|^2} \int \Psi(x) \rho_0(x) dx + \int \Psi^*(y) \rho_1(y) dy$$

$$\mathcal{W}_2^2(\rho_0 - \epsilon \eta, \rho_1) \geq \int \Psi(x) (\rho_0 - \epsilon \eta) dx + \int \Psi^*(y) \rho_1(y) dy$$

$$\frac{\mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_1) - \mathcal{W}_2^2(\rho_0, \rho_1)}{\epsilon} \geq \int_{\Omega} \Psi \eta$$

Proof

Kantorovich potential :

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \max_{\Psi, \Psi(x) + \Psi^*(y) \leq |x-y|^2} \int \Psi(x) \rho_0(x) dx + \int \Psi^*(y) \rho_1(y) dy$$

$$\mathcal{W}_2^2(\rho_0 - \epsilon \eta, \rho_1) \geq \int \Psi(x) (\rho_0 - \epsilon \eta) dx + \int \Psi^*(y) \rho_1(y) dy$$

$$\frac{\mathcal{W}_2^2(\rho_0 + \epsilon \eta, \rho_1) - \mathcal{W}_2^2(\rho_0, \rho_1)}{\epsilon} \geq \int_{\Omega} \Psi \eta$$

$$\frac{\mathcal{W}_2^2(\rho_0, \rho_1) - \mathcal{W}_2^2(\rho_0 - \epsilon \eta, \rho_1)}{\epsilon} \leq \int_{\Omega} \Psi \eta$$

Back to the computation of the gradient of $\mathcal{J}_{\mathcal{W}}$

Let's finish the computation of the gradient of $\mathcal{J}_{\mathcal{W}}$:

$$\mathcal{J}_{\mathcal{W}}(\rho_0 + \epsilon\eta) =$$

$$\sum_i \mathcal{W}_2^2\left(\mathcal{M}_i\rho_0 + \epsilon \frac{\partial \mathcal{M}_i}{\partial \rho_0}(\rho_0) \cdot \eta + o(\epsilon), \rho_i^{\text{obs}}\right) + \omega_b \mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_0^b)$$

Back to the computation of the gradient of $\mathcal{J}_{\mathcal{W}}$

Let's finish the computation of the gradient of $\mathcal{J}_{\mathcal{W}}$:

$$\begin{aligned} \mathcal{J}_{\mathcal{W}}(\rho_0 + \epsilon\eta) &= \\ \sum_i \mathcal{W}_2^2\left(\mathcal{M}_i\rho_0 + \epsilon\frac{\partial\mathcal{M}_i}{\partial\rho_0}(\rho_0)\cdot\eta + o(\epsilon), \rho_i^{\text{obs}}\right) &+ \omega_b\mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_0^b) \\ &= \mathcal{J}_{\mathcal{W}}(\rho_0) + \epsilon\sum_i \left\langle \frac{\partial\mathcal{M}_i}{\partial\rho_0}(\rho_0)\cdot\eta, \Phi^i \right\rangle_2 + \epsilon\langle\eta, \omega_b\Phi^b\rangle_2 + o(\epsilon) \end{aligned}$$

Back to the computation of the gradient of $\mathcal{J}_{\mathcal{W}}$

Let's finish the computation of the gradient of $\mathcal{J}_{\mathcal{W}}$:

$$\begin{aligned} \mathcal{J}_{\mathcal{W}}(\rho_0 + \epsilon\eta) &= \\ \sum_i \mathcal{W}_2^2\left(\mathcal{M}_i\rho_0 + \epsilon \frac{\partial \mathcal{M}_i}{\partial \rho_0}(\rho_0) \cdot \eta + o(\epsilon), \rho_i^{\text{obs}}\right) &+ \omega_b \mathcal{W}_2^2(\rho_0 + \epsilon\eta, \rho_0^b) \\ &= \mathcal{J}_{\mathcal{W}}(\rho_0) + \epsilon \sum_i \left\langle \frac{\partial \mathcal{M}_i}{\partial \rho_0}(\rho_0) \cdot \eta, \Phi^i \right\rangle_2 + \epsilon \langle \eta, \omega_b \Phi^b \rangle_2 + o(\epsilon) \end{aligned}$$

$$\mathcal{J}_{\mathcal{W}}(\rho_0 + \epsilon\eta) = \mathcal{J}_{\mathcal{W}}(\rho_0) + \epsilon \left\langle \eta, \sum_i \frac{\partial \mathcal{M}_i^*}{\partial \rho_0} \cdot \Phi^i + \omega_b \Phi^b \right\rangle_2 + o(\epsilon)$$

Gradient of \mathcal{W}_2

- Thus, the \mathcal{L}_2 gradient is

$$\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \frac{\partial \mathcal{M}_i^*}{\partial \rho_0} \cdot \Phi^i + \omega_b \Phi^b$$

Gradient of \mathcal{W}_2

- Thus, the \mathcal{L}_2 gradient is

$$\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \frac{\partial \mathcal{M}_i^*}{\partial \rho_0} \cdot \Phi^i + \omega_b \Phi^b$$

- And the \mathcal{W}_2 gradient is

$$\text{grad}_{\mathcal{W}} \mathcal{J}_{\mathcal{W}}(\rho_0) = -\text{div}(\rho_0 \nabla[\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0)])$$

Gradient descent

- Recall that

$$\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \frac{\partial \mathcal{M}_i^*}{\partial \rho_0} \cdot \Phi^i + \omega_b \Phi^b$$

$$\text{grad}_{\mathcal{W}} \mathcal{J}_{\mathcal{W}}(\rho_0) = -\text{div}(\rho_0 \nabla[\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0)])$$

Gradient descent

- Recall that

$$\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \frac{\partial \mathcal{M}_i^*}{\partial \rho_0} \cdot \Phi^i + \omega_b \Phi^b$$

$$\text{grad}_{\mathcal{W}} \mathcal{J}_{\mathcal{W}}(\rho_0) = -\text{div}(\rho_0 \nabla [\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0)])$$

- Descent gradient writes

$$\rho_0^{n+1} = \rho_0^n - \alpha^n \text{grad} \mathcal{J}_{\mathcal{W}}(\rho_0^n)$$

Gradient descent

- Recall that

$$\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \frac{\partial \mathcal{M}_i^*}{\partial \rho_0} \cdot \Phi^i + \omega_b \Phi^b$$

$$\text{grad}_{\mathcal{W}} \mathcal{J}_{\mathcal{W}}(\rho_0) = -\text{div}(\rho_0 \nabla [\text{grad}_2 \mathcal{J}_{\mathcal{W}}(\rho_0)])$$

- Descent gradient writes

$$\rho_0^{n+1} = \rho_0^n - \alpha^n \text{grad} \mathcal{J}_{\mathcal{W}}(\rho_0^n)$$

- But which gradient ?

When the cost function is written with the help of the Wasserstein distance, it's better to take the Wasserstein gradient

Note on the Wasserstein gradient descent

- The gradient descent writes

$$\rho_0^{n+1} = \rho_0^n + \alpha^n \operatorname{div}(\rho_0^n \nabla \Phi^n)$$

Note on the Wasserstein gradient descent

- The gradient descent writes

$$\rho_0^{n+1} = \rho_0^n + \alpha^n \operatorname{div}(\rho_0^n \nabla \Phi^n)$$

- Instead, we rather write it as

$$\rho_0^{n+1} = (\operatorname{Id} - \alpha^n \nabla \Phi^n) \# \rho_0^n$$

Note on the Wasserstein gradient descent

- The gradient descent writes

$$\rho_0^{n+1} = \rho_0^n + \alpha^n \operatorname{div}(\rho_0^n \nabla \Phi^n)$$

- Instead, we rather write it as

$$\rho_0^{n+1} = (\operatorname{Id} - \alpha^n \nabla \Phi^n) \# \rho_0^n$$

- Or, equivalently, $\rho_0^{n+1} = \rho(t = \alpha^n)$ with

$$\begin{cases} \rho(t=0) = \rho_0^n, & \Phi(t=0) = -\Phi^n \\ \partial_t \rho + \operatorname{div}(\rho \nabla \Phi) = 0 \\ \partial_t \Phi + \frac{|\nabla \Phi|^2}{2} = 0 \end{cases}$$

- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

Twin experiment

- Let the model be the continuity equation

$$\begin{cases} \rho(t=0) = \rho_0 \\ \partial_t \rho + \operatorname{div}(\rho u) = 0 \end{cases}$$

with $u(t, x)$ known and will not be looked for.

Twin experiment

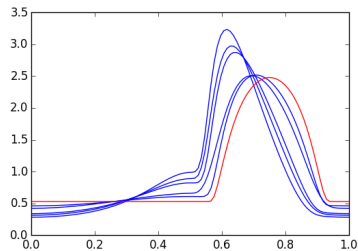
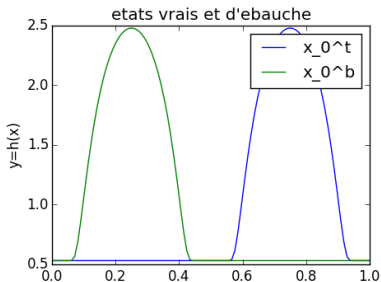
- Let the model be the continuity equation

$$\begin{cases} \rho(t=0) = \rho_0 \\ \partial_t \rho + \operatorname{div}(\rho u) = 0 \end{cases}$$

with $u(t, x)$ known and will not be looked for.

- Observations are $\rho_i^{\text{obs}} := \mathcal{M}_i \rho_0^t$.

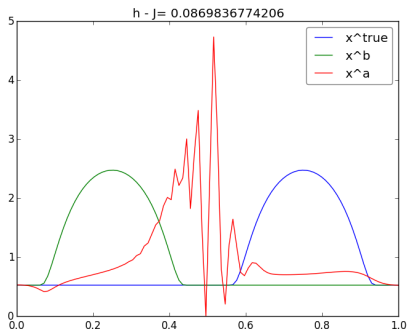
The background term has some error displacements.



Observations

$$\mathcal{J}_{\mathcal{W}}(\rho_0) = \sum_i \mathcal{W}_2^2(\mathcal{M}_i \rho_0, \underbrace{\rho_i^{\text{obs}}}_{\mathcal{M}_i \rho_0^t}) + \omega_b \mathcal{W}_2^2(\rho_0, \rho_0^b)$$

Results...

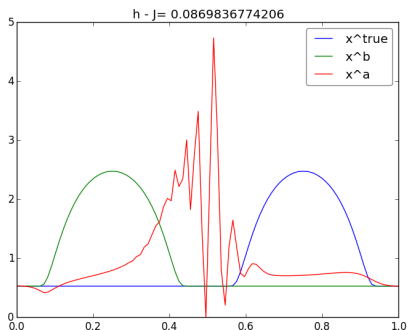


Oscillations + descent is stuck because of the 0 value at $x = 0.5$.

- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

White noise "invisible" to Wasserstein distance

$$\mathcal{W}_2^2(\rho^n, \rho) \rightarrow 0 \Rightarrow \rho^n \rightarrow \rho$$



Solutions to oscillations

Sources of oscillations :

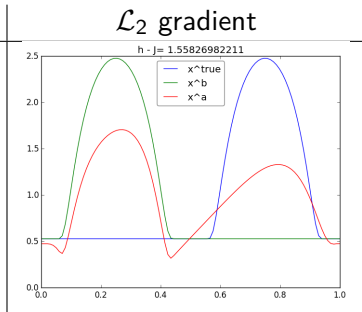
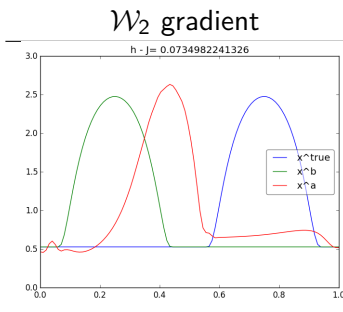
- Wasserstein distance is not sensible to oscillations.
- The continuous equation scheme creates oscillations as we solve

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho \nabla \Phi) = 0 \\ \partial_t \Phi + \frac{|\nabla \Phi|^2}{2} = 0 \end{cases} \quad (1)$$

Solutions :

- Add a small \mathcal{L}_2 term in the cost function : $\epsilon \|\rho_0\|_2^2$
- Add a small artificial viscosity term in (1).

In general case...



- 1 Optimal transport
 - Definitions
 - Properties of the Wasserstein distance
- 2 Data assimilation
 - Definition
 - Application of OT to DA
- 3 Minimizing \mathcal{J}_W , numerically
 - Computing W_2^2
 - Descent gradient of \mathcal{J}_W
 - Wasserstein inner product
- 4 Results
 - First results
 - Solve oscillations
 - Perspectives

Use better descent algorithms

- Gradient descent is quite slow.
- Use BFGS or conjugate gradient (CG) method instead. Need an adaptation because of the inner product.

When densities are not of same mass

If ρ_0 and ρ_1 are not of same mass, the WF_δ distance introduced in [1], [2] (Wasserstein-Fisher-Rao distance)

$$WF_\delta = \inf_{\partial_t \rho + \operatorname{div}(\rho v) = \rho \Psi} \iint \rho |v|^2 + \delta \rho |\Psi|^2$$

which has an underlying inner product (thus, perfect for minimizing !)

$$\langle \eta, \eta' \rangle_\delta := \int \rho \nabla \Phi \cdot \nabla \Phi' + \delta \int \rho \Phi \Phi' \quad \text{with } \eta = -\operatorname{div}(\rho \nabla \Phi) + \delta \rho \Phi$$

[1] : KONDRATYEV, Stanislav, MONSAINGEON, Léonard, et VOROTNIKOV, Dmitry. A new optimal transport distance on the space of finite Radon measures. arXiv preprint arXiv :1505.07746, 2015.

[2] : CHIZAT, Lenaic, PEYRÉ, Gabriel, SCHMITZER, Bernhard, et al. Unbalanced optimal transport : geometry and Kantorovich formulation. arXiv preprint arXiv :1508.05216, 2015.

When there are unobserved variables

How to write $\mathcal{J}_{\mathcal{W}}(\rho_0, u_0)$?

When there are unobserved variables

How to write $\mathcal{J}_{\mathcal{W}}(\rho_0, u_0)$?

$$\mathcal{J}_{\mathcal{W}}(\rho_0, u_0) = \mathcal{W}_2^2(\rho_i(\rho_0, u_0), \rho^i) + \omega_\rho \mathcal{W}_2^2(\rho_0, \rho_0^b) + \omega_u \mathcal{W}_2^2(u_0, u_0^b)$$

When there are unobserved variables

How to write $\mathcal{J}_{\mathcal{W}}(\rho_0, u_0)$?

$$\mathcal{J}_{\mathcal{W}}(\rho_0, u_0) = \mathcal{W}_2^2(\rho_i(\rho_0, u_0), \rho^i) + \omega_\rho \mathcal{W}_2^2(\rho_0, \rho_0^b) + \omega_u \mathcal{W}_2^2(u_0, u_0^b)$$

$$\mathcal{W}_2^2(\rho_0, u_0; \rho_0^b, u_0^b) = \inf_{\substack{\partial_t \rho + \operatorname{div}(\rho v) = 0 \\ \partial_t u + \operatorname{div}(\rho w) = 0}} \iint_{\Omega} [\rho |v|^2 + \alpha \rho |w|^2] \, dt dx$$

Thank you for your attention !