ESAIM: Mathematical Modelling and Numerical Analysis

NUMERICAL RESOLUTION OF AN "UNBALANCED" MASS TRANSPORT PROBLEM

Jean-David Benamou[1]

Abstract. We introduce a modification of the Monge–Kantorovitch problem of exponent 2 which accommodates non balanced initial and final densities. The augmented Lagrangian numerical method introduced in [6] is adapted to this "unbalanced" problem. We illustrate the usability of this method on an idealized error estimation problem in meteorology.

1. INTRODUCTION

From Monge's theory "des déblais et des remblais" in the late 18th century [35] to present days, a considerable amount of work in mass transport theory has been gathered [10,12,19,22,24,28,30,33,36,37,39]. Apart from its use in different theoretical studies [2,11,13–15,38]) potential "industrial" applications also emerge, for instance in medical imaging, astronomy, meteorology etc. [9, 21, 23, 26, 29, 31, 32, 34]. The design of efficient numerical methods is therefore already an issue for the applied mathematicians [1, 3, 4, 6–8, 17, 26, 27].

A major obstacle to the actual use of the Monge–Kantorovitch mass transfer problem (MKP) – that yields the Wasserstein distance between density functions – is the total mass balance constraint on the data of the problem: The solution of the MKP provides an optimal transfer map between two (initial and final) prescribed density functions with identical L^1-norm. In other words the classical MKP provides a transfer map preserving the total mass. One rarely faces "real" balanced data, therefore using the MKP on unbalanced data either requires some normalization or a modification of the MKP. Different normalizations have been tested in [34] and it seems that such a process unavoidably modify the mass distribution and therefore the transfer map itself. If the balance default is caused by noise or other non significant phenomena, a normalization of the data can seriously impair the relevance of the MKP solution.

We therefore focus on the second option and propose a modification of the the MKP which is similar to the penalization technique presented in [7]. We here adapt the augmented Lagrangian technique of [6] to the time-dependent and penalized formulation of the MKP and obtain a robust numerical method for unbalanced data.

Section 2 presents the mathematical model, Section 3 the augmented Lagrangian numerical method. We illustrate the method in Section 4 on a particular application related to an error estimation problem in meteorology.

2. Mixed L^2-Wasserstein distance

2.1. The Monge–Kantorovitch problem

Let us first recall the mono-phasic "classical" formulation of the Monge–Kantorovitch problem (MKP).

Given two bounded, non-negative measurable functions ρ_0 and ρ_T with same L^1 norm (same mass) and compact support in a bounded subset D of $\mathbb{R}^d$, find an application M which realizes the transport from ρ_0 to ρ_T in the following sense: for all Borel set A, M satisfies

$$\int_{M^{-1}(A)} \rho_0(x)\mathrm{d}x = \int_A \rho_T(x)\mathrm{d}x, \tag{1}$$

and achieves the minimal cost

$$\int_D |x - M(x)|^2 \rho_0(x)\mathrm{d}x. \tag{2}$$

2.2. The Wasserstein distance

We recall here the time-dependent formulation of the Wasserstein distance (for details see [6] or [25]). We are given two density functions $\rho_0(x) > 0$ and $\rho_1(x) > 0$ of $x \in D$, a bounded open set in $\mathbb{R}^d$. We further assume that they are bounded and normalized – we will say "balanced":

$$\int \rho_0(x)\mathrm{d}x = \int \rho_1(x)\mathrm{d}x = 1. \tag{3}$$

We fix a time interval $[0, T]$ ($T = 1$ in the sequel) and consider all possible time-dependent, density et velocity fields, $\rho(t, x) > 0$, $v(t, x) \in \mathbb{R}^d$, satisfying the continuity equation

$$\partial_t \rho + \nabla.(\rho v) = 0, \quad v \cdot \vec{n} = 0 \ \text{ on } \partial D \tag{4}$$

for $0 < t < T$ and $x \in D$, $\vec{n}$ is the outward normal to ∂D. The boundary condition express that the mass cannot flow out of the zone of interest D.

Initial and final conditions

$$\rho(0, .) = \rho_0, \ \ \rho(T, .) = \rho_1 \tag{5}$$

are prescribed for equation (4). Then, the square of the Wasserstein distance (also called L^2-Kantorovitch distance – see [19] for a review) is given by

$$d_{\mathrm{was}}(\rho_0, \rho_1)^2 = \inf_{\rho, v} \left\{ \frac{1}{T} \int \int_0^T \rho(t, x)|v(t, x)|^2 \mathrm{d}x \, \mathrm{d}t \right\}, \tag{6}$$

among all (ρ, v) satisfying (4) and (5). The minimization problem (6) can be shown to be convex with linear constraints in the variable $(\rho, m = \rho v)$. Indeed using these new variables:

$$d_{\mathrm{was}}(\rho_0, \rho_1)^2 = \inf_{\rho, m} \left\{ \frac{1}{T} \int \int_0^T \frac{|m|^2}{2\rho} \mathrm{d}x \, \mathrm{d}t \right\} \tag{7}$$

and (4) transforms in

$$\partial_t \rho + \nabla.(m) = 0 \ \ m \cdot \vec{n} = 0 \ \text{ on } \partial D. \tag{8}$$

The convexity with respect to (ρ, m) of the integrand in (7) follows from formula (22). The solution is unique and and an augmented Lagrangian numerical technique has been proposed in [6] for its computation.

Remark 2.1. As already mentioned in this previous paper, a striking property of this distance is its smoothness with respect to translations. Let us assume for instance that ρ_1 is compactly supported in $\mathbb{R}^d$ and obtained from ρ_0 through a translation τ_μ

$$\rho_1(x) = \rho_0(x - \mu), \tag{9}$$

for $\mu \in \mathbb{R}^d$. Then it can be checked that the density is analytically given by

$$\rho(t, x) = \rho_0\left(x - \frac{t}{T}\mu\right) \tag{10}$$

and the optimization problem (4) simply yields a quadratic dependence of the distance with respect to the translation:

$$d_{\text{was}}(\rho_0, \rho_1)^2 = |\mu|^2 \int \rho_0(x)\,\mathrm{d}x. \tag{11}$$

Remark 2.2. It is possible to express the L^2-distance using a similar time dependent formulation. We indeed remark that

$$d_{L^2}(\rho_0, \rho_1)^2 = \inf_{\rho, v}\left\{\frac{1}{T}\int\int_0^T |\partial_t \rho(t,x)|^2 \mathrm{d}x\,\mathrm{d}t\right\}, \tag{12}$$

with (ρ, v) again subject to the constraints (2) and (3). The optimization problem does not depend anymore on v and constraint (2) is just mentioned here by analogy with Section 1.1. We again have to consider a convex minimization problem. The cost function appearing on the right hand side of (11) can be differentiated. The optimality condition simply express that the optimal $\rho(t, x)$ satisfies (3) and

$$\partial_{tt}^2 \rho = 0 \tag{13}$$

and is therefore simply given by the time interpolation formula

$$\rho(t, x) = \frac{\rho_1(x) - \rho_0(x)}{T}t + \rho_0(x). \tag{14}$$

Replacing $\partial_t \rho$ in (11), we finally obtain the claimed L^2-distance:

$$d_{L^2}(\rho_0, \rho_1)^2 = \int |\rho_1(x) - \rho_0(x)|^2 \mathrm{d}x.$$

2.3. The mixed relaxed distance

It is clear (see [7,16,34]) that the behavior with respect to translation and dilation makes, in some applications, the choice of the distance a sensitive issue and the Wasserstein distance may be a good choice. However the constraint of mass balance (3) is rarely satisfied.

We proposed in [7] a penalization formulation for unbalanced mass data. A classical technique in optimal control for dealing with two point (initial and final) boundary conditions is to relax one of these, for instance the final condition $\rho(T, .) = \rho_1$ and enforce it through a penalization term added to the cost functional. More precisely instead of (6), we used the relaxed MKP

$$d_{\text{relax}}(\rho_0, \rho_1)^2 = \inf_m\left\{T\int\int_0^T \frac{|m|^2}{2\rho}\mathrm{d}x\mathrm{d}t + \frac{\gamma}{2}\int_D |\rho(T, x) - \rho_1(x)|^2 \mathrm{d}x\right\} \tag{15}$$

where γ is a positive penalty parameter. Note that ρ is not anymore a minimization variable. It is a state variable defined by ρ_0 and (8). The final density $\rho(T, x)$ is necessarily balanced with ρ_0 because it is a solution to the conservation law (4) but not necessarily with ρ_1 now decoupled from the constraints. Existence and

uniqueness of a solution for problem (15) was formally established in [7] where we use a conjugate gradient method. This numerical method turns out to be sensitive to the implicit constraint $\rho > 0$ which is necessary to the convex characterization of the integrand (formula (22)). One goal of this paper is to adapt the more robust augmented Lagrangian method of [6] to a slightly different modification of the MKP involving the L^2-distance.

2.4. **The mixed distance**

In this paper we propose to work on unbalanced data by considering the mixed Wasserstein/L^2-distance in the following sense: given two possibly unbalanced densities ρ_0 and ρ_1, find $\tilde{\rho}_1$ – the closest density to ρ_1 in the L^2-sense – which minimizes the Wasserstein distance $d_{\text{wass}}(\rho_0, \tilde{\rho}_1)$. It can be formulated as

$$\inf_{\tilde{\rho}_1} \left\{ d_{\text{wass}}(\rho_0, \tilde{\rho}_1)^2 + \frac{\gamma}{2} d_{L^2}(\tilde{\rho}_1, \rho_1)^2 \right\} \tag{16}$$

where γ is a positive parameter balancing the two distances.

Remark that this problem is conceptually very close to (15) even though we here keep the minimization with respect to the whole ρ field. It can be reformulated as

$$d_{\text{mix}}(\rho_0, \rho_1)^2 = \inf_{\rho, m, \tilde{\rho}_1} \left\{ T \int \int_0^T \frac{|m|^2}{2\rho} \mathrm{d}x\,\mathrm{d}t + \frac{\gamma}{2} \int |\tilde{\rho}_1(x) - \rho_1(x)|^2 \mathrm{d}x \right\} \tag{17}$$

with the additional constraint

$$\rho(T, .) = \tilde{\rho}_1 \tag{18}$$

together with the usual constraint $\rho(T, .) = \rho_0$ and continuity equation (8) to be satisfied.

One can observe that constraint (18) just indicates that the minimization variables $\rho(T, .)$ and $\tilde{\rho}_1$ have to be the same. Eliminating $\tilde{\rho}_1$ reduces (17) to (15). We here focus on formulation (17) and solve it by adapting the augmented Lagrangian method of [6].

3. THE AUGMENTED LAGRANGIAN NUMERICAL METHOD

3.1. **The Lagrangian**

The computational bounded domain is denoted $D \in \mathbb{R}^d$ and the following notations will be used:

- ∇_x is the spatial gradient in $\mathbb{R}^d$;
- $\nabla_{t,x} = \{\partial_t, \nabla_x\}$ is the time-space gradient in $\mathbb{R}^{d+1}$;
- $\Delta_{t,x} = \partial_t^2 + \Delta_x$ is the time-space Laplacian;
- for two vectors in $\mathbb{R} \times \mathbb{R}^d$, a, b and a', b', $\{a, b\}.\{a', b'\} = aa' + b.b'$ denotes the inner product.

The MKP can be written as a saddle-point problem by introducing a space-time dependent Lagrange multiplier $\phi(t, x)$ for constraints (4) and (5).

The Lagrangian is given by:

$$L(\phi, \rho, m, \tilde{\rho}_1) = \int_0^T \int_D \frac{|m|^2}{2\rho} - \partial_t \phi \rho - \nabla_x \phi.m \,\mathrm{d}x\,\mathrm{d}t - \int_D \phi(0, .)\rho_0 - \phi(T, .)\tilde{\rho}_1 \mathrm{d}x + \frac{\gamma}{2} \int_D |\tilde{\rho}_1 - \rho_1|^2 \mathrm{d}x \tag{19}$$

where the terms involving ϕ come from the integration by part of (4) and using boundary conditions (5–18).

The MKP is equivalent to the saddle-point problem:

$$\inf_{\rho, m, \tilde{\rho}_1} \sup_\phi L(\phi, \rho, m, \tilde{\rho}_1). \tag{20}$$

The (formal) optimality conditions for this problem are:

$$
\begin{cases}
\partial_t \phi + \dfrac{|m|^2}{2\rho^2} = 0, \quad \dfrac{m}{\rho} = \nabla_x \phi, \\[2mm]
\partial_t \rho + \nabla_x . m = 0, \\[2mm]
\rho(0,.) = \rho_0, \\[2mm]
\rho(T,.) = \tilde{\rho}_1, \\[2mm]
\phi(T,.) = \gamma(\rho_1 - \tilde{\rho}_1).
\end{cases}
\tag{21}
$$

Notice that, if m is eliminated, we recover the equations of a pressureless potential flow. If $\tilde{\rho}_1$ is eliminated we get a "classical" direct/adjoint optimality system (in the optimal control theory).

3.2. A new Lagrangian

We continue to adapt the augmented Lagrangian strategy of [6] and build a new Lagrangian for which the methods of [20] can be applied. Let us first observe that, for positive ρ (which becomes an implicit constraint), we have, pointwise in time and space,

$$
\frac{|m(t,x)|^2}{2\rho(t,x)} = \sup_{\{a,b\} \in K} [a(t,x)\rho(t,x) + b(t,x).m(t,x)]
\tag{22}
$$

where

$$
K = \left\{ \{a,b\} : \mathbb{R} \times D \to \mathbb{R} \times \mathbb{R}^d, \ \ s.\,t. \ \ a + \frac{|b|^2}{2} \leq 0, \ \ \forall (t,x) \in \mathbb{R} \times D \right\}.
\tag{23}
$$

Notice that the left-hand side of (22) becomes infinite whenever ρ vanishes or ρ is positive and m vanishes. It is also important to notice that the optimal b $(= \frac{m}{\rho})$ above has to satisfy $b \cdot \vec{n} = 0$ on ∂D because of the boundary conditions on m. We can thus also restrain K to

$$
K = \left\{ \{a,b\} : \mathbb{R} \times D \to \mathbb{R} \times \mathbb{R}^d, \ \ s.\,t. \ \ a + \frac{|b|^2}{2} \leq 0. \ \forall (t,x) \in \mathbb{R} \times D \ \ \text{and} \ \ b \cdot \vec{n} = 0, \ \ \forall (t,x) \in \mathbb{R} \times \partial D \right\}.
\tag{24}
$$

Using formula (22) the non-linear integrand can be replaced by a linear function that has to be optimized with respect to the new variable (a,b). This term will eventually be combined to produce a new linear constraint that can advantageously be "augmented" as proposed in [20] (see below).

Consequently it is natural to adopt the same strategy on the non-linear L^2-norm penalization. We use the Legendre transform of $x \to \frac{x^2}{2}$ to dualize the penalization term:

$$
\frac{|\tilde{\rho}_1 - \rho_1|^2}{2} = \sup_c \left[c(\tilde{\rho}_1 - \rho_1) - \frac{c^2}{2} \right]
\tag{25}
$$

where c is a function on D and (25) holds pointwise.

We follow the terminology of [6] and use the following variables and notations:

$$
\mu = \{\rho, m\} \quad q = \{a,b\}, \quad \langle \mu, q \rangle = \int_0^T \int_D \mu.q,
\tag{26}
$$

$$
F(q) = 0 \text{ if } q \in K, \quad +\infty \text{ else}, \quad G(\phi) = \int_D \phi(0,.)\rho_0 \, dx.
\tag{27}
$$

We also need the new functional:

$$H(c) = \gamma \int_D c\rho_1 + \frac{c^2}{2} \, \mathrm{d}x \tag{28}$$

and for clarity will set $\psi = \phi(T, .)$ a redundant notation for the Lagrange multiplier at time T.

After some calculations and changing the inf/sup saddle point problem into an sup/inf, we rewrite problem (20) as:

$$\sup_{\mu, \tilde{\rho}_1} \inf_{\phi, \psi, q, c} \left[F(q) + G(\phi) + H(c) + \langle \mu, \nabla_{t,x}\phi - q \rangle - \int_D [\tilde{\rho}_1(\psi + \gamma c)] \, \mathrm{d}x \right]. \tag{29}$$

Notice that μ and $\tilde{\rho}_1$ are now Lagrange multipliers of linear constraints which will be "augmented" in the next section. To comply fully with the hypothesis of [20], the integrand of H should be linear (like G) – it is here non linear (but nicely quadratic). Also, the function F should be coercive and is not. In [6], we used a coercive ϵ perturbation of F: $F_\epsilon(q) = F + \epsilon|q|^2$. In practice such a modification does not seem necessary and recent theoretical developments [25] back this observation.

3.3. The augmented Lagrangian

As mentioned above, in the new saddle-point problem (29) we now consider $\mu = \{\rho, m\}$ and $\tilde{\rho}_1$ as the Lagrange multipliers of constraints acting on ϕ, q, ψ ($= \phi(T, .)$) and c, namely

$$\nabla_{t,x}\phi - q = 0, \quad \psi + \gamma c = 0. \tag{30}$$

Thus, following [20] we define the "augmented" Lagrangian:

$$L_r(\phi, q, c, \mu, \tilde{\rho}_1) = F(q) + G(\phi) + H(c) + \langle \mu, \nabla_{t,x}\phi - q \rangle - \int_D [\tilde{\rho}_1(\psi + \gamma c)] \, \mathrm{d}x$$

$$+ \frac{r_1}{2} \langle \nabla_{t,x}\phi - q, \nabla_{t,x}\phi - q \rangle + \frac{r_2}{2} \int_D [(\psi + \gamma c)^2] \, \mathrm{d}x \tag{31}$$

where r_1, r_2 are positive parameters which are *a priori* chosen identical in the sequel of this paper. We recall that ψ is a shorthand notation for $\phi(T, .)$. The corresponding saddle-point problem

$$\sup_{\mu, \tilde{\rho}_1} \inf_{\phi, q, c} L_r(\phi, q, c, \mu, \tilde{\rho}_1) \tag{32}$$

clearly has the same solution as (20).

3.4. The algorithm

A simple algorithm of [20], called ALG2, based on relaxations of the Uzawa algorithm is now used to solve the problem. We get a three step iterative method which constructs a sequence $(\phi^n, \psi^n, q^n, c^n, \mu^n, \tilde{\rho}_1{}^n)$ expected to converge to the saddle-point.

ALG2:
- $(\phi^{n-1}, q^{n-1}, \mu^n, \tilde{\rho}_1{}^n)$ are given.
- Step A: Find ϕ^n such that:

$$L_r(\phi^n, q^{n-1}, c^{n-1}, \mu^n, \tilde{\rho}_1{}^n) \le L_r(\phi, q^{n-1}, c^{n-1}, \mu^n, \tilde{\rho}_1{}^n), \quad \forall \phi. \tag{33}$$

  Then set

$$\psi^n = \phi^n(T, .).$$

- Step B: Find q^n, c^n such that:

$$L_r(\phi^n, q^n, c^n, \mu^n, \tilde{\rho}_1{}^n) \le L_r(\phi^n, q, c, \mu^n, \tilde{\rho}_1{}^n), \quad \forall q, c. \tag{34}$$

- Step C: Do the gradient descent

$$\mu^{n+1} = \mu^n + r_1(\nabla_{t,x}\phi^n - q^n)$$
$$\tilde{\rho_1}^{n+1} = \tilde{\rho_1}^n - r_2(\psi^n + \gamma c^n)$$

(35)

where, apparently, the custom is to use r_1, r_2 the parameter of the augmented constraints as the gradient steps (hence maybe the need to pick up different rs for the two constraints).
- Go back to step A.

Step A and B simply are a relaxation method for the minimization part of the saddle-point problem. Step C is a gradient step for the dual problem.

3.5. The practical algorithm

We now discuss each of step of the abstract algorithm for our particular Lagrangian.

Step A

To solve Step A, we need to differentiate L_r with respect to ϕ and get for ϕ^n:

$$G(\phi) + r_1 < \nabla_{t,x}\phi^n - q^{n-1}, \nabla_{t,x}\phi > + < \mu^n, \nabla_{t,x}\phi > + r_2 \int_D [(\phi^n(T,.) + \gamma c^{n-1}(.))\phi(T,.)] = 0, \ \forall \phi.$$

After integrating by part in space and time and using the various boundary conditions we see that this is the variational formulation of the space-time Laplace equation

$$-r_1\Delta_{t,x}\phi^n = \nabla_{t,x}.\left(\mu^n - r_1 q^{n-1}\right)$$

(36)

with boundary conditions:

$$r_1\partial_t\phi^n(0,.) = \rho_0 - \rho^n(0,.) + r_1 a^{n-1}(0,.),$$

(37)

$$r_1\partial_t\phi^n(T,.) + r_2\phi^n(T,.) = r_1 a^{n-1}(T,.) - r_2\gamma c^{n-1}(.)$$

(38)

and

$$r_1\nabla_x\phi^n(0,.) \cdot \vec{n} = r_1 b^{n-1}(0,.) \cdot \vec{n}.$$

(39)

Notice that (37) is a weak way of enforcing the initial density conditions. The strong way is to force $\rho^n(0,.) = \rho_0$.

We use a publicly available fortran package (HS3CRT by R. Sweet) that uses a FFT technique to solve this problem on a finite difference staggered grid.

Step B

We cannot differentiate L_r with respect to q and q^n is simply obtained by solving

$$\inf_q [F(q) + \frac{r_1}{2} < \nabla_{t,x}\phi^n - q, \nabla_{t,x}\phi^n - q > + < \mu^n, \nabla_{t,x}\phi^n - q >],$$

which is equivalent to:

$$inf_{q\in K} < \nabla_{t,x}\phi^n + \frac{\mu^n}{r_1} - q, \nabla_{t,x}\phi^n + \frac{\mu^n}{r_1} - q > .$$

It is important to notice that this minimization can be performed pointwise in space and time. Indeed, let us set:

$$p^n(t,x) = \{\alpha^n(t,x), \beta^n(t,x)\} = \nabla_{t,x}\phi^n(t,x) + \frac{\mu^n(t,x)}{r_1}.$$

Then $q^n(t,x) = \{a^n(t,x), b^n(t,x)\}$ is obtained by solving in (a,b):

$$\inf_{(a,b)\in K} \left\{ (a - \alpha^n(t,x))^2 + |b - \beta^n(t,x)|^2, \quad a + \frac{|b|^2}{2} \leq 0 \right\}.$$

This is turns out to be a simple one dimensional projection problem which can be computed analytically or using Newton's method. The projection of the discretized variables turns out to be not that simple as the α and β variable have to be defined on two different staggered grid. This problem is analyzed in detail in [25].

Minimizing L_r with respect to c is easy (quadratic) and gives pointwise in space

$$\gamma c^n = \frac{\gamma(\tilde{\rho_1}^n - \rho_1) - r_2\psi^n}{(1 + r_2\gamma)}.$$

Again ψ^n $(=\phi^n(T,.))$ cannot be discretized on the same grid points as $\tilde{\rho_1}^n$ and c^n. They are staggered and distant of half a time step dt. In practice we make a linear approximation using the final time boundary condition:

$$\psi^n = \frac{\phi^n(T,.) + \frac{dt}{2}(a^{n-1}(T,.) - \gamma c^{n-1})}{1 + \frac{dt}{2}}.$$

Step C

Step C is simply the pointwise update – a gradient step for the dual problems:

$$\mu^{n+1} = \mu^n + r_1(\nabla_{t,x}\phi^n - q^n),$$

$$\tilde{\rho}_1^{n+1} = \tilde{\rho}_1^n - r_2(\psi^n + \gamma c^n).$$

As discussed in [6] each iteration of this algorithm costs an optimal $O(N \log N)$ operations where N is the number of grid points in time and space.

3.6. Convergences

ALG2 convergence

We first check that the ALG2 iteration process converges. We let the algorithm run over a large (1000) number of iterations and observe four criteria: the Wasserstein and L^2-distance components of the mixed distance

$$d_{\text{wass}}(\rho_0, \tilde{\rho}_1(x))^2 = \int \int_0^T \frac{|m|^2}{2\rho} dx \, dt, \quad d_{L^2}(\rho_1, \tilde{\rho}_1(x))^2 = \int |\tilde{\rho}_1(x) - \rho_1(x)|^2 dx \tag{40}$$

and the duality errors given as L^2-norms of the gradients with respect to the dual variables μ and $\tilde{\rho}_1$

$$\text{duality error } 1 = \left(\int \int_0^T \|\nabla_{t,x}\phi^n - q^n\|^2 dx \, dt\right)^{\frac{1}{2}}, \quad \text{duality error } 2 = \left(\int |\psi^n + \gamma c^n|^2 dx\right)^{\frac{1}{2}}. \tag{41}$$

We take $D =]-1,1[\times]-1,1[$, $x = (x_1, x_2) \in \mathbb{R}^2$, $T = 1$, $\gamma = 1$ and $r_1 = r_2 = 1$. We discretize the three dimensional space-time domain using a $32 \times 32 \times 32$ grid. We show on Figures 1 and 2 the evolution of the four criteria above for two tests:

Test 1. We use balanced initial and final densities, two Gaussian functions centered respectively at $(0.7, 0.7)$ and $(-0.7, -0.7)$:

$$\rho_0(x) = e^{-30\left((x_1-0.7)^2+(x_2-0.7)^2\right)}, \quad \rho_1(x) = e^{-30\left((x_1+0.7)^2+(x_2+0.7)^2\right)}.$$

Test 2. We perturb the final density so that the data in now not balanced:

$$\rho_0(x) = e^{-30\left\{(x_1-0.7)^2+(x_2-0.7)^2\right\}}, \quad \rho_1(x) = e^{-30\left\{(x_1+0.7)^2+(x_2+0.7)^2\right\}} + 0.2e^{-40\left(x_1^2+x_2^2\right)}.$$

Both (balanced and unbalanced) tests indicate a convergence thus confirming the convexity of our mixed functional. Note that as in the pure Wasserstein case the deterioration of the convergence rate indicates a rather
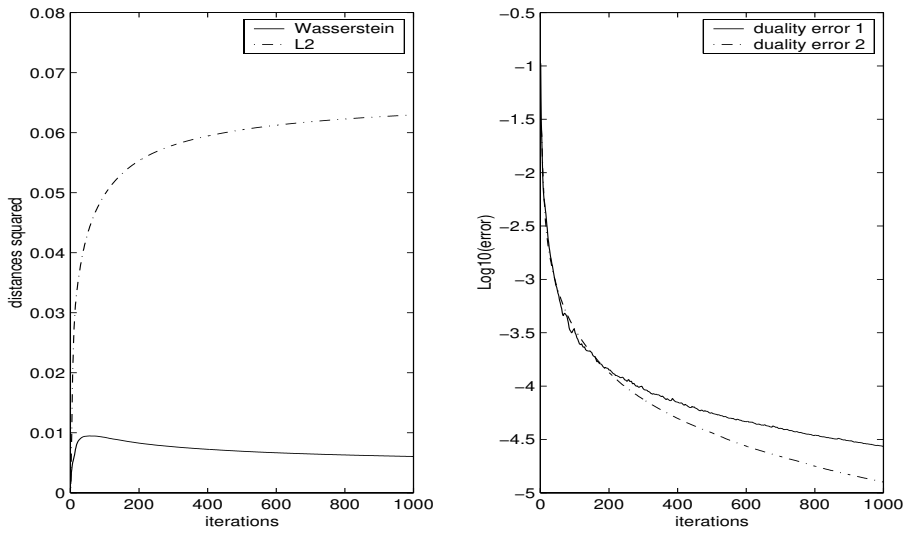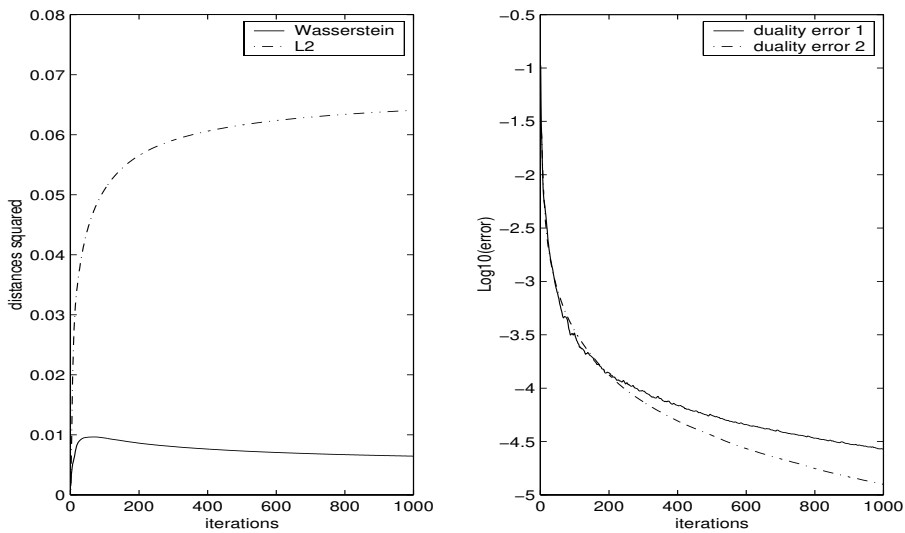
FIGURE 1. Test 1.

FIGURE 2. Test 2.

flat cost function. We emphasize that we did not study the impact of the r_1, r_2 parameters – in particular we always set $r_1 = r_2$.

γ convergence

We here investigate the impact of the parameter γ on the produced solution. In particular we want to check, for balanced data, the expected convergence towards respectively the Wasserstein and L^2-distances. The results of the table below concern the balanced densities of **Test 1** above and use the same numerical parameters except for γ.

Let us be more precise: we observe, after convergence of the ALG2 algorithm, the values of the Wasserstein and L^2-components of the mixed distance. When γ increases, the L^2-penalization forces the final computed

density to match the given final density and the Wasserstein component of the distance (second line in the table) approaches the correct value of Wasserstein distance. Thanks to the imposed boundary conditions, our test behaves as if we translate a Gaussian in all space (and not on the torus). The exact solution is given by (see (11))

$$d_{\mathrm{was}}(\rho_0, \rho_1)^2 = |\mu|^2 \int \rho_0(x)\,\mathrm{d}x$$

where μ is the translation distance. In our case $\mu^2 = 2(1.4)^2 = 3.92$ and a quadrature formula gives $\int \rho_0(x) = 0.102745727$ so that a reliable estimate of the Wasserstein distance is

$$d_{\mathrm{was}}(\rho_0, \rho_1)^2 = 0.4027$$

which is in good accordance with the value obtained for $\gamma = 100$ in the table.

When γ decreases, the impact of the penalization is weaker and the mass transfer activity decreases. The initial density tend to remains stationary and therefore the L^2-component of the distance (third line in the table) goes to the L^2-norm of the difference between initial and final prescribed densities. In our **Test 1** case

$$\int |\rho_1(x) - \rho_0(x)|^2\,\mathrm{d}x = 0.1046$$

which matches the L^2-component of the mixed distance for $\nu = 0.01$.

γ	0.01	0.1	1	10	100
$\int \int_0^T \frac{|m|^2}{2\rho}\,\mathrm{d}x\,\mathrm{d}t$	1.7×10^{-7}	4.7×10^{-5}	0.0060	0.3754	0.4099
$\int |\tilde{\rho}_1(x) - \rho_1(x)|^2\,\mathrm{d}x$	0.1042	0.0925	0.0629	0.0001	2.2×10^{-8}

Another way to understand the γ dependence of the problem is to look at the time evolution of the density $\rho(., x)$. Figures 3 and 5 show the level curves of $\rho(t, x)$ for different time steps and for the two extreme values of γ. When $\gamma = 100$ the final density is forced to match ρ_1 and then the mass transport minimizes the Wasserstein criterion (translation). The elliptical shape of the Gaussian is probably caused by the missing mass – the initial and final Gaussian are partially supported outside the domain. Notice (Fig. 4) that the convergence is slower for large γ. When $\gamma = 0.01$ the L^2- component is not so important and then we minimize a Wasserstein criterion with no imposed final density. The optimal solution does not transfer mass at all and small noise appears (Fig. 5).

4. APPLICATION: DRIFT ERRORS BETWEEN UNBALANCED PERTURBED DATA

4.1. **Motivation**

In this section we experiment our algorithm in view of a particular target application suggested by Mike Cullen (see also [16, 18]). It concerns the problem of the quantitative comparison of weather forecast with the actual observed weather in view of assessing the quality of the forecasting technique. This problem is therefore typically the computation of a distance between two density functions. One represents the forecast of a physical quantity (the rain density for instance) at a fixed date and time and the other is the actual observed weather variable at same date and time.

There are apparently two sources of errors that respectively exclude L^2 and Wasserstein distances. The first type of error is generally a drift: the forecast is reasonably accurate both in shape and intensity but, maybe because of errors in winds, slightly translated compared to the actual observed data. The L^2-distance here is
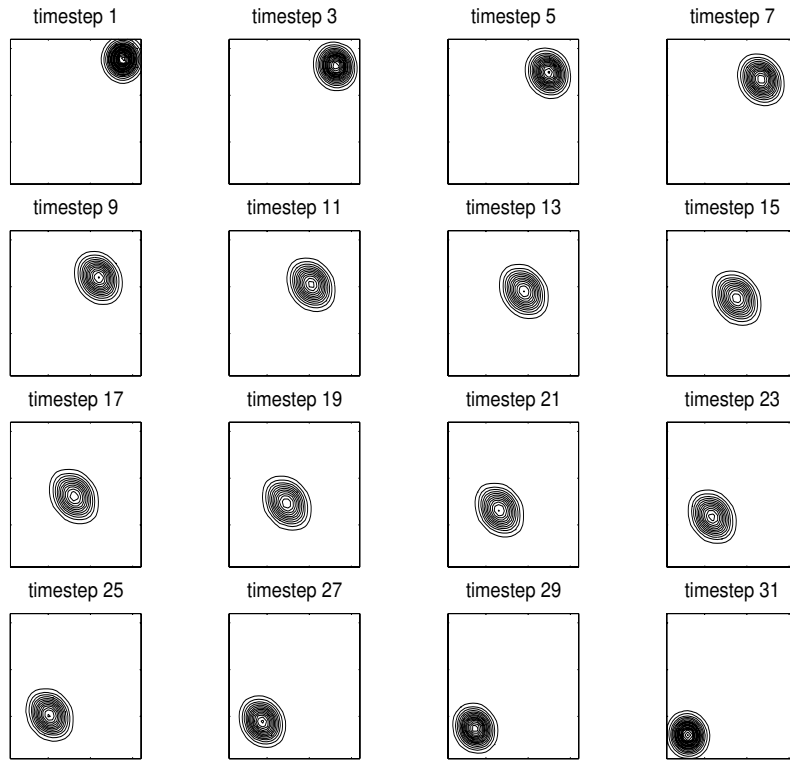
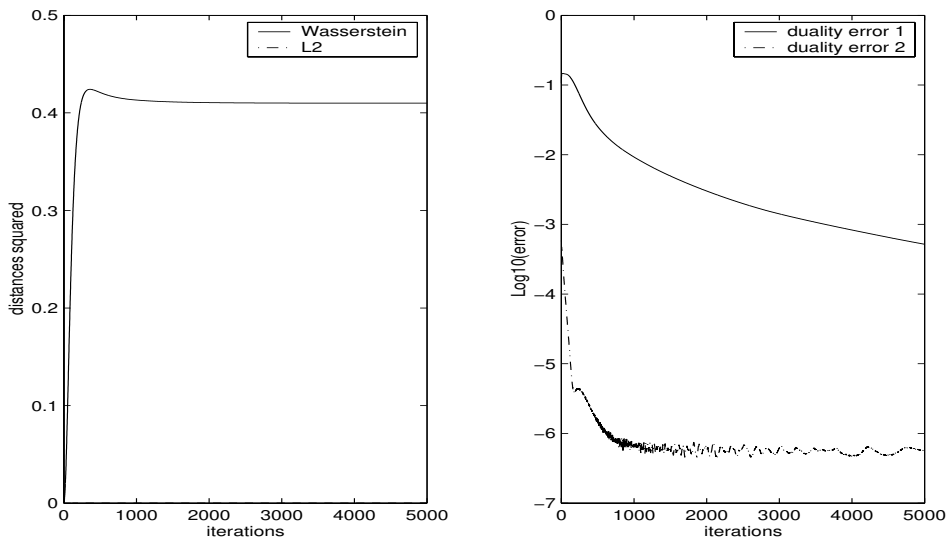FIGURE 3. Level curves of ρ for $\gamma = 100$.
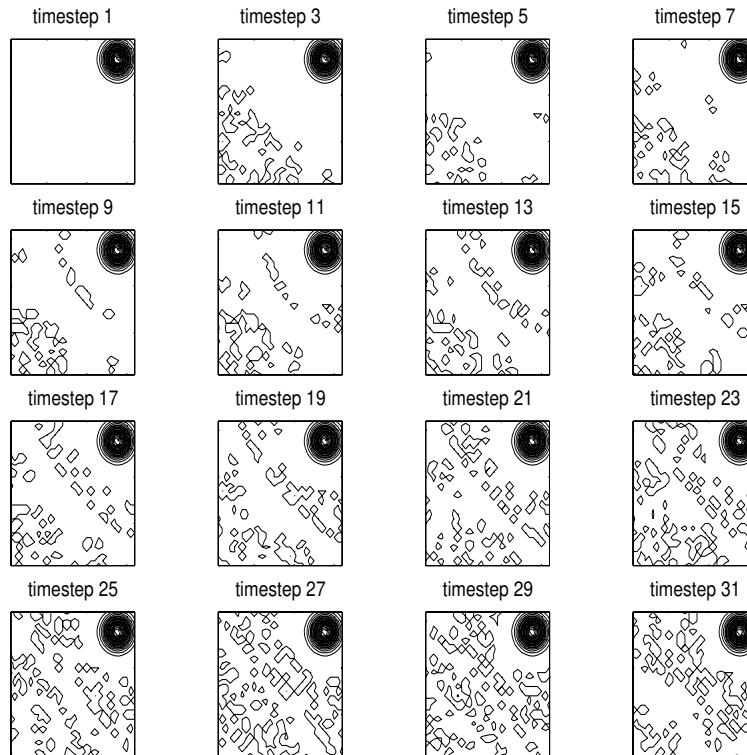

FIGURE 4. Convergence.

FIGURE 5. Level curves of ρ for $\gamma = 0.01$.

a poor choice because it is very sensitive to translations and cannot take into account the "good" part of the forecast: shape and intensity. As discussed in [7], Wasserstein distance is here an excellent choice (provided data are balanced). Indeed, formula (11) indicates that this distance is very smooth with respect to translations.

The second source of errors originates in local perturbations away from the main zones of activity which cannot be forecasted either because of the physics limitations of the model or insufficient observed initial data that is fed into it. Here the Wasserstein distance is not a good choice because of this additional mass – we expect the main tracked zone to roughly conserve the mass. So using a normalization process to balance data unavoidably causes additional mass transport which is charged to the Wasserstein distance evaluation while having nothing in common with the first "translation" source of errors. Conversely the L^2-distance is locally good at the perturbation. As we assume the forecast to predict low or nil intensities in these zones the local L^2-norm difference is a relevant measure of this particular failure of the forecasting technique.

As we can see, a subtle combination of both distance is needed. One would like to measure – using an automatic procedure – the drift error using the Wasserstein distance and the local perturbation errors using the L^2-norm. We explain in Sections 4.2 and 4.3 how, under certain limitations, our mixed distance can achieve this goal on idealized test cases.

4.2. Balancing (1D)

We first focus on a simple 1D problem to try and estimate the values of γ that suits the needs described in the previous section. Let (Fig. 6)
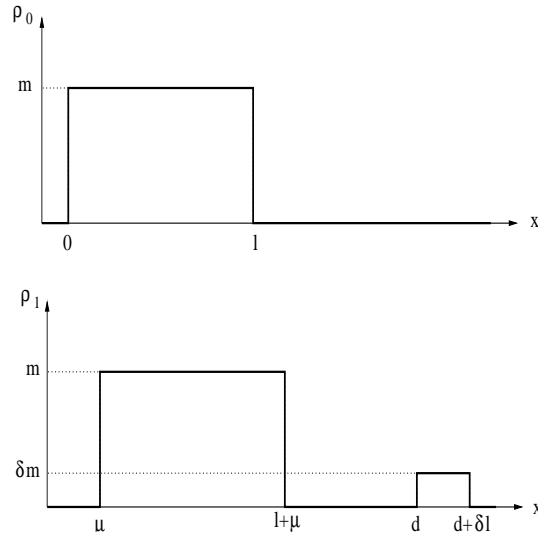
$$\rho_0(x) = m \, \Xi_{]0,l[}(x)$$

FIGURE 6. Initial and final densities.

be the characteristic function of segment $]0, l[$ times a constant m. In accordance with the above described application, we set (all parameters are positive):

$$\rho_1(x) = m \, \Xi_{]\mu, l+\mu[}(x) + \delta\rho(x)$$

where μ is a (small) translation parameter and $\delta\rho$ a perturbation (ρ_0 and ρ_1 are not balanced) taken as

$$\delta\rho(x) = \delta m \, \Xi_{]d, d+\delta l[}(x)$$

where δm and δl are small with respect to m and l; and d is such that the supports are disjoint:

$$]d, d + \delta l[\cap]0, l[= \emptyset, \qquad]d, d + \delta l[\cap]\mu, l + \mu[= \emptyset.$$

We recall that the optimization problem we solve can be written

$$\inf_{\tilde\rho_1} C_{\rho_0, \rho_1}(\tilde\rho_1) \tag{42}$$

where the cost function is defined by

$$C_{\rho_0, \rho_1}(\tilde\rho_1) = d_{\text{wass}}(\rho_0, \tilde\rho_1)^2 + \frac{\gamma}{2} \, d_{L^2}(\rho_0, \tilde\rho_1)^2. \tag{43}$$

We now evaluate this cost function for three different values of $\tilde\rho_1$ – balanced with ρ_0 – corresponding both to the pathological and desired behaviors of the algorithm:

a) No transport

As explained in Section 3.6, γ may be so low that no mass at all (or very little) is transported. It corresponds – in the limit case $\gamma = 0$ – to

$$\tilde\rho_1 = \rho_0.$$

The cost function then yields

$$
\begin{aligned}
C_a &= C_{\rho_0,\rho_1}(\rho_0) \\
&= d_{\text{wass}}(\rho_0,\rho_0)^2 + \frac{\gamma}{2} \left\| m\, \Xi_{]\mu,l+\mu[} - m\, \Xi_{]0,l[} \right\|^2 + \frac{\gamma}{2} \left\| \Xi_{]d,d+\delta l[} \right\|^2 \\
&= 0 + \gamma\, \mu\, m^2 + \frac{\gamma}{2}\, \delta l\, \delta m^2.
\end{aligned}
$$

b) Partial transport

Conversely, when γ is very large and the data not balanced the optimization problem can override the mass transport criterion and allocate mass to minimize the L^2-component even if the transport cost is higher. The extreme case looks like

$$
\tilde{\rho}_1 = \delta\rho + \frac{m\,l - \delta l\,\delta m}{l}\, \Xi_{]\mu,l+\mu[}.
$$

Some mass has been transfered to fill in the perturbation $\delta\rho$ and there is not enough left to translate exactly the main characteristic function, hence the parameter $\frac{m\,l - \delta l}{l}$ in front of the second term in the above formula which adjusts the height of the characteristic function to ensure the mass balance of $\tilde{\rho}_1$ with ρ_0.

We now compute the cost function

$$
\begin{aligned}
C_b &= d_{\text{wass}}(\rho_0,\tilde{\rho}_1)^2 + \frac{\gamma}{2} \left\| m\, \Xi_{]\mu,l+\mu[} - \frac{m\,l - \delta l}{l}\, \Xi_{]\mu,l+\mu[} \right\|^2 \\
&\simeq d^2\, \delta l\, \delta m + \mu^2 (m\,l - \delta l\,\delta m) + \frac{\gamma}{2}\, \frac{\delta l^2\, \delta m^2}{l}
\end{aligned}
$$

where we approximate the Wasserstein part by two separate translation cost.

c) Perfect transport

We finally arrive at the desired solution. We wish γ to be such that the Wasserstein part of the cost function picks up the small mass μ translation and the L^2 part the perturbation $\delta\rho$. In other word we want the optimal $\tilde{\rho}_1$ to be

$$
\tilde{\rho}_1 = m\, \Xi_{]\mu,l+\mu[}.
$$

In which case the cost function gives

$$
C_c = \mu^2\, m\,l + \frac{\gamma}{2}\, \delta l\, \delta m^2.
$$

The rule on γ

We can now derive a heuristic rule on γ that satisfies our objectives. Definitely, necessary conditions to get the desired $\tilde{\rho}_1$ are

$$
C_c < C_a, \qquad C_c < C_b.
$$

These two conditions translate into

$$
\frac{\mu\,l}{m} < \gamma < 2\, \frac{l}{\delta m\,(l - \delta l)}\, \left(d^2 - \mu^2\right). \tag{44}
$$

We illustrate numerically this section using $l = 1$, $m = 10$, $\mu = 0.1$, $d = 1.7$, $\delta l = 0.1$, $\delta m = 1$. With these data (44) becomes

$$
0.01 < \gamma < 6.4.
$$

All 1D computations are performed using 32 points in time and 128 in space. As the mixed distance cost function – like Wasserstein distance's – is convex but very flat, a few hundred iterations are sufficient to get a good estimate of the cost function.
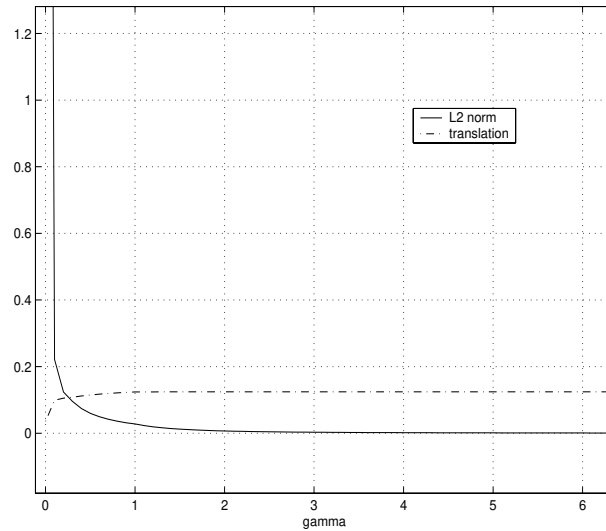
FIGURE 7. γ dependence.

Figure 7 shows the numerically estimated translation (from formula (11))

$$\mu = \sqrt{\frac{d_{\text{was}}(\rho_0, \tilde{\rho_1})^2}{\int \rho_0(x)\, \mathrm{d}x}} \tag{45}$$

and L^2-norm $d_{l^2}(\rho_1, \tilde{\rho_1})^2$ as a function of γ. Notice these are non-linear but monotone functions. One can also observe behavior analyzed above. The "good" value of γ, *i.e.* for which we recover the correct values of the translation and L^2-error (here both 0.1) is approximately 0.25.

We show in Figure 8 both ρ_1 the prescribed final density – not balanced with ρ_0 – and $\tilde{\rho_1}$ the optimal final density solution of (16) – that has the same mass as $\tilde{\rho}(_0)$.

4.3. γ optimization

Our algorithm allows to compute amongst other things the mapping

$$\gamma \to d_{l^2}(\rho_1, \tilde{\rho_1})^2 \tag{46}$$

where $\tilde{\rho_1}$ is the optimal final density. As we explained in the previous section we would like to "tune" γ such that $\tilde{\rho_1} = 0$ in the perturbed zone and $\tilde{\rho_1} = \rho_0(x - \mu)$ in the zone where translation occurs. After convergence we therefore should get

$$d_{l^2}(\rho_1, \tilde{\rho_1}) = \int |\delta\rho(x)|^2 \,\mathrm{d}x.$$

In our synthetic test example we of course know this quantity ($\int |\delta\rho(x)|^2 \,\mathrm{d}x = 0.1$) and we can invert numerically the function defined by the mapping (46). We used a simple dichotomy algorithm which gives after 10 iterations

$$\gamma = 0.2783, \;\; d_{L^2}(\rho_1, \tilde{\rho_1}) = 0.0989, \;\; \mu = \sqrt{\frac{d_{\text{was}}(\rho_0, \tilde{\rho_1})^2}{\int \rho_0(x)\, \mathrm{d}x}} = 0.1080.$$

We show in Figure 8 the associated optimal density $\tilde{\rho}(_1)$ (down left) compared to ρ_1 (top right) the prescribed final density. Most of mass has correctly been translated only in the main region. One can note the presence of residual mass on the left of the main characteristic function and we have no explanation for that.
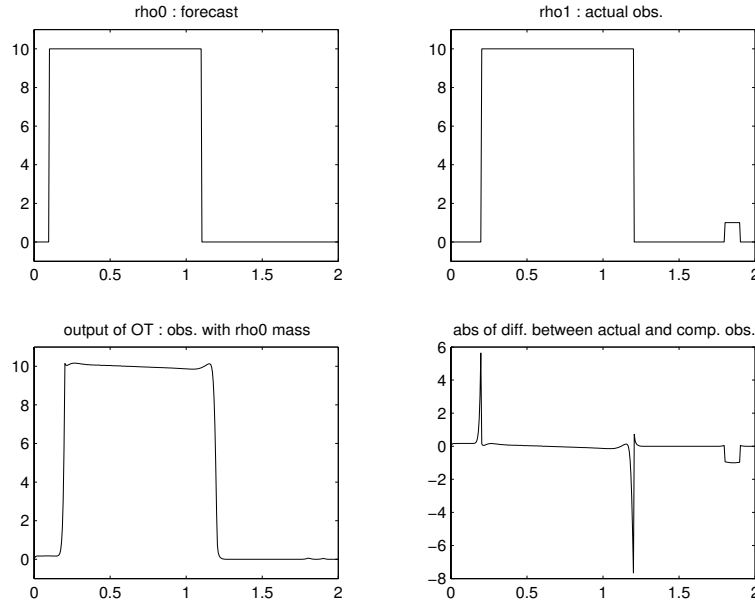
FIGURE 8. Row wise: ρ_0, ρ_1, $\tilde{\rho_1}$, $\tilde{\rho_1}$, $\tilde{\rho_1} - \rho_1$.

Remark that in practice of course $\delta\rho$ is not exactly known but if the region where the translation occurs is roughly known then $\rho_1 - \rho_0$ outside this domain should be a good guess for $\delta\rho$.

4.4. 2D results

We now present results which indicate that the numerical strategy applied in the previous section is also relevant in 2D. Our test case is defined by

$$\rho_0(x) = 10\,e^{-30\left((x_1-0.1)^2+(x_2-0.1)^2\right)}, \quad \rho_1(x) = 10\,e^{-30\left(x_1^2+x_2^2\right)} + \delta\rho(x)$$

where the theoretical translation is therefore $\mu = 0.1\sqrt{2}$ and the perturbation

$$\delta\rho(x) = e^{-80((x_1-0.7)^2+(x_2-0.7)^2)} + e^{-80((x_1+0.7)^2+(x_2-0.7)^2)} + e^{-80((x_1-0.7)^2+(x_2+0.7)^2)} + e^{-80((x_1+0.7)^2+(x_2+0.7)^2)}$$

has separate support and L^2-norm

$$\int |\delta\rho(x)|^2 dx \simeq 0.078539774.$$

We use 16 points in time and a 64×64 grid in space.

The dichotomy algorithm described in Section 4.3 allows to find an optimal $\gamma = 0.2159$ for which

$$d_{l^2}(\rho_1, \tilde{\rho_1}) = 0.0784689113, \quad \mu = 0.13.$$

We show in Figure 9 the prescribe initial and and final densities, the optimal density $\tilde{\rho_1}$ and the absolute value of the difference between ρ_1 and $\tilde{\rho_1}$.

5. CONCLUSION AND PROSPECTS

All our numerical tests have demonstrated the robustness and efficiency of the method – it converges – even for discontinuous and compactly supported data (such as characteristic functions).
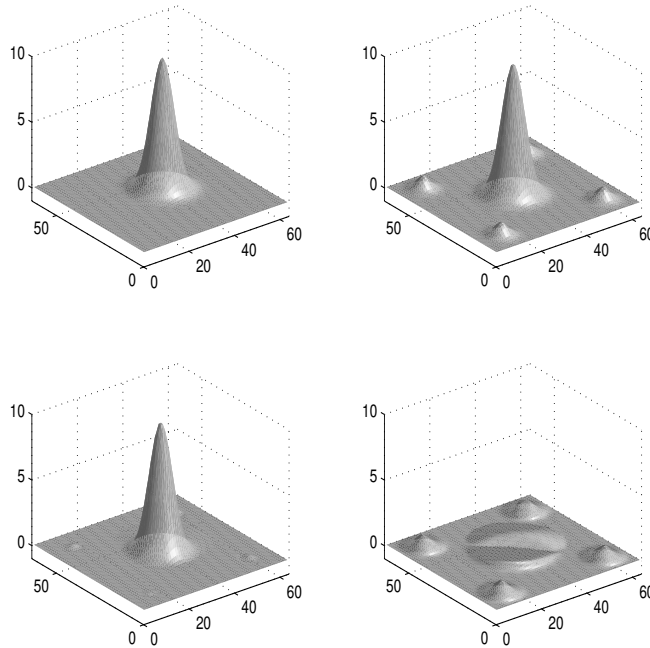
FIGURE 9. From left top line-wise: ρ_0, ρ_1, $\tilde{\rho}_1$ $|\tilde{\rho}_1 - \rho_1|$.

We would like to point out that we incorporated more realistic non-permeable boundary conditions (our previous papers and code only worked for periodic boundary conditions in space).

The method seems well adapted to the depicted application at least in the idealized cases where the optimization of γ allows to roughly separate the Wasserstein translation phenomena from the L^2-error estimate.

We did not investigate a number of open questions such as:

- The impact of parameters r_1 and r_2 on the speed of convergence.
- The impact of the space discretization. In particular a coarse discretization causes a loss of the mass conservation property for large γ. Small Spurious oscillation may also appear at sharp density discontinuities and may – inconsistently with the model – provoke the apparition of negative mass.
- Incorporate dilations in the idealized test case.
- The generation of negative mass (Rem. 4.1).
- The underestimation of the translation in the 2D case.
- As usual time is an artificial dimension and the question of time discretization is still non investigated. We just tried to compute the same 2D test case using only 4 time steps instead of 16, we found

$$d_{l^2}(\rho_1, \tilde{\rho}_1) = 0.0784689113, \ \ \mu = 0.126951993$$

which differs only by a few percents. This observation backs up the idea that a coarse time discretization is sufficient to represent the fairly simple time behavior of the problem (the optimal trajectories – the characteristics of equation (4) – are straight lines in the pure Wasserstein case).

References

[1] M. Balinski, A competitive (dual) simplex method for the assignment problem. *Math. Program.* **34** (1986) 125–141.

[2] F. Barthe, On a reverse form of the Brascamp-Lieb inequality. *Invent. Math.* **134** (1998) 335–361.

[3] J.-D. Benamou, A domain decomposition method for the polar factorization of vector valued mappings. *SIAM J. Numer. Anal.* **32** (1995) 1808–1838.

[4] J.D. Benamou and Y. Brenier, Numerical resolution on a massively parallel computer of a test problem in meteorology using a domain decomposition algorithm, in *First European conference in computational fluid dynamics*. North Holland (1992).

[5] J.D. Benamou and Y. Brenier, Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampère/transport problem. *SIAM J. Appl. Math.* **58** (1998) 1450–1461.

[6] J.D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.* **84** (2000) 375–393.

[7] J.D. Benamou and Y. Brenier, *Mixed L^2/Wasserstein Optimal Mapping Between Prescribed Densities Functions* (submitted).

[8] J.D. Benamou, Y. Brenier and K. Guittet, Numerical resolution of a multiphasic optimal mass transport problem. *Tech. Report INRIA RR-4022*.

[9] G. Boucjitte, G. Buttazzo and P. Seppechere, Shape Optimization Solutions via Monge-Kantorovich. *C. R. Acad. Sci. Paris Sér. I* **324** (1997) 1185–1191.

[10] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** (1991) 375–417.

[11] Y. Brenier, Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Comm. Pure Appl. Math.* **52** (1999) 411–452.

[12] Y. Brenier, *Extended Monge-Kantorovich theory.* CIME 2001 lecture.

[13] L.A. Caffarelli, Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.* **45** (1992) 1141–1151.

[14] L.A. Caffarelli, Boundary regularity of maps with convex potentials. II. *Ann. of Math.* **144** (1996) 3, 453–496.

[15] M.J.P. Cullen, Solution to a model of a front forced by deformation. *Q. J. R. Met. Soc.* **109** (1983) 565–573.

[16] M.J.P. Cullen, private communication.

[17] M.J.P. Cullen and R.J. Purser, An extended Lagrangian theory of semigeostrophic frontogenesis. *J. Atmopheric Sci.* **41** (1984) 1477–1497.

[18] R.J. Douglas, *Decomposition of weather forecast error using rearrangements of functions.* (Preprint.)

[19] L.C. Evans, *Partial differential equations and Monge-Kantorovich mass transfer.* Lecture notes.

[20] M. Fortin and R. Glowinski, *Augmented Lagrangian methods. Applications to the numerical solution of boundary value problems.* North-Holland Publishing Co. Studies in Mathematics and its Applications 15 (1983) 340.

[21] U. Frisch *et al.*, Back to the early Universe by optimal mass transportation. *Nature* **417** (2002) 260–262.

[22] W. Gangbo and R.J. McCann, The geometry of optimal transportation. *Acta Math.* **177** (1996) 113–161.

[23] W. Gangbo and R.J. McCann, Shape recognition via Wasserstein distance. *Quart. Appl. Math.* **58** (2000) 705–737.

[24] K. Guittet, On the time-continuous mass transport problem and its approximation by augmented Lagrangian techniques. *SIAM J. Numer. Anal.* **41** (2003) 382–399.

[25] K. Guittet, Ph.D. dissertation (2002).

[26] S. Haker, A. Tannenbaum and R. Kikinis, Mass preserving mapping and image registration. *MICCAI* (2001) 120–127.

[27] R. Jonker and A. Volgenant, A shortest augmenting path algorithm for dense and sparse linear assignment problem. *Computing* **38** (1987) 325–340.

[28] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** (1998) 1–17.

[29] T. Kaijser, Computing the Kantorovich distance for images. *J. Math. Imaging Vision* **9** (1998) 173–198.

[30] L.V. Kantorovich, On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **37** (1942) 199–201.

[31] D. Kinderlehrer and N. Walkington, Approximation of Parabolic Equations based upon a Wasserstein metric. *ESAIM: M2AN* **33** (1999) 837–852.

[32] S.A. Kochengin and V.I. Oliker, Determination of reflector surfaces from near-field scattering data. *Inverse Problems* **13** (1997) 363–373.

[33] R.J. McCann, Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.* **11** (2001) 589–608.

[34] R. Menozzi, *Utilisation de la distance de Wasserstein et application sismique.* Rapport IUP Génie Mathématique et Informatique, Université Paris IX-Dauphine.

[35] G. Monge, *Mémoire sur la théorie des déblais et des remblais.* Mem. Acad. Sci. Paris (1781).

[36] F. Otto, The geometry of dissipative evolution equation: the porous medium equation. *Comm. Partial Differential Equations* **26** (2001) 101–174.

[37] S.T. Rachev and L. Rüschendorf, Mass transportation problems, in *Theory, Probability and its Applications, Vol. I.* Springer-Verlag, New York (1998) 508.

[38] A. Shnirelman, Generalized fluid flows, their approximation and applications. *Geom. Funct. Anal.* **4** (1994) 586–620.

[39] C. Villani, *Topics in mass transport.* Lecture notes (2000).