Journal of Computational Physics 230 (2011) 3430–3442

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

A lagrangian scheme for the solution of the optimal mass transfer problem

Angelo Iollo *, Damiano Lombardi

Institut de Mathématiques de Bordeaux UMR 5251 Université Bordeaux 1 and INRIA Bordeaux-Sud Ouest, équipe-projet MC2, 351, Cours de la Libération, Talence 33405, France

ARTICLE INFO

Article history:
Received 20 May 2010
Received in revised form 24 November 2010
Accepted 25 January 2011
Available online 4 February 2011

Keywords:
Monge–Kantorovich problem
Optimal transport
Lagrangian methods

ABSTRACT

A lagrangian method to numerically solve the L^2 optimal mass transfer problem is presented. The initial and final density distributions are approximated by finite mass particles having a gaussian kernel. Mass conservation and the Hamilton–Jacobi equation for the potential are identically satisfied by constant mass transport along straight lines. The scheme is described in the context of existing methods to solve the problem and a set of numerical examples including applications to medical imagery are presented.

1. Introduction

The optimal mass transfer problem, also known as the Monge–Kantorovich problem (see [1,2]), consists in finding a plan to transport a certain quantity of mass from a starting configuration to a final one, minimizing a given cost functional. Optimal transport theory has recently been revived and developed providing theoretical tools for the analysis of phenomena such as geophysical flows [3], nonlinear electrodynamics [3–5], collapsing sand piles [6], crowd motion [7]. An exhaustive overview of the theory can be found in [8–10].

In this work we focus on the numerical solution of the L^2 optimal mass transfer problem in $\mathbb{R}^d$, where d is the number of space dimensions. Let $\rho_0(\xi)$ and $\rho_1(x)$ be two non-negative scalar (density) functions with compact support Ω_0 and Ω_1, respectively. We assume that

$$\int_{\Omega_0} \rho_0(\xi)d\xi = \int_{\Omega_1} \rho_1(x)dx = 1. \tag{1}$$

Let $x = X(\xi)$ be a smooth one-to-one map taking Ω_0 onto Ω_1 that verifies the jacobian equation

$$det(\nabla_\xi X)\rho_1(X(\xi)) = \rho_0(\xi). \tag{2}$$

As a consequence, we have that $\forall \Omega \subseteq \Omega_0$

$$\int_{\Omega} \rho_0(\xi)d\xi = \int_{X(\Omega)} \rho_1(x)dx. \tag{3}$$

The jacobian equation (2) has many admissible solutions. Among all these mappings, the objective of this paper is to describe a lagrangian method to find $X^*(\xi)$ such that

* Corresponding author.
E-mail addresses: angelo.iollo@math.u-bordeaux1.fr (A. Iollo), damiano.lombardi@math.u-bordeaux1.fr (D. Lombardi).

0021-9991/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2011.01.037

$$\int_{\Omega_0} \rho_0(\xi)\|X^*(\xi)-\xi\|^2 d\xi \leqslant \int_{\Omega_0} \rho_0(\xi)\|X(\xi)-\xi\|^2 d\xi \tag{4}$$

for all smooth one-to-one mappings $X(\xi)$. This functional measures the cost of the mass transport by a weighted square distance function. Other classes of optimal transport problems can be defined by introducing different norms instead of the above. We concentrate on the L^2 case because of its links with continuum mechanics [3] and since the solution of this problem finds applications in oceanography [11], shape optimization [12], computer vision [13] and image processing [14].

Mainly two classes of methods to solve this problem in realistic applications were proposed. One idea (see [15]) is to look for a mapping between the initial and final condition by solving an appropriate partial differential equation up to steady state. The computational cost of this approach is that of finding the asymptotic solution of d transport equations. The main drawback is that, apart from accuracy, the convergence rate to the asymptotic solution may be poor. A different path is followed in [16], where a time-like variable is introduced and the space–time mapping between the initial and final mass distribution is found by a saddle point method that requires the solution of a Poisson problem in space and time at each iteration. The merit of the latter formulation is to show the links between the least action principle and the optimal mass transfer problem. From the computational view point, however, the time-like variable introduces additional unknowns to be solved for. In the next sections we will summarize the formulation at the base of these two approaches in order to introduce an alternative solution method where no partial differential equations are numerically solved to approximate the optimal map.

2. The Angenent–Haker–Tannenbaum (AHT) gradient flow

Let us recall a basic theoretical result on the L^2 optimal mass transfer problem (see [17,10,18]): there is a unique optimal map $X^*(\xi)$ characterized as the unique map transferring $\rho_0(\xi)$ to $\rho_1(x)$ which can be written as the gradient of some convex function $\Psi(\xi)$:

$$X^*(\xi) = \nabla_\xi \Psi(\xi); \tag{5}$$

in other words if we find a map that can be expressed as in the equation above and that satisfies Eq. (2), than this is the optimal map. Such a consideration is at the base of the AHT method [15].

In the following we recast AHT method in a continuum mechanics framework. Let $x = \mathcal{X}(\xi,t)$ be a smooth one-to-one mapping such that $\mathcal{X}(\xi,0) = X_i(\xi)$ and let the initial map $X_i(\xi)$ satisfy the jacobian equation (2). The objective is to make this initial map evolve toward the optimal map by a gradient method, without altering the distribution $\rho_1(x)$.

Our plan is therefore to compute the variation of the cost functional

$$I = \int_{\Omega_0} \rho_0(\xi)\|\mathcal{X}(\xi,t)-\xi\|^2 d\xi, \tag{6}$$

with respect to t. To do so, we need some preliminary steps. We introduce the inverse mapping $\xi = Y(x,t)$, that, for given value of the parameter t and space coordinate $x \in \Omega_1$ retrieves the corresponding $\xi \in \Omega_0$. In other words $x = \mathcal{X}(Y(x,t),t)$ and hence

$$\nabla_\xi \mathcal{X} = (\nabla_x Y)^{-1}, \tag{7}$$

$$\partial_t Y + U\cdot\nabla_x Y = 0, \tag{8}$$

where $U(x,t) = \partial_t \mathcal{X}(\xi,t)$. If $\mathcal{X}(\xi,t)$ has to take Ω_0 onto Ω_1, $\forall t \in \mathbb{R}^+$, then $\forall x \in \partial\Omega_1$ and $\forall t \in \mathbb{R}^+, U(x,t)\cdot n = 0$, where n is the normal to $\partial\Omega_1$. As an initial condition for the inverse map we take $Y(x,0) = Y_i(\xi)$ with $Y_i(x) = X_i^{-1}(x)$. Let also assume that $x = \mathcal{X}(\xi,t)$ is mass preserving $\forall t \in \mathbb{R}^+$ so that

$$det(\nabla_x Y(x,t))\rho_0(Y(x,t)) = \rho(x), \tag{9}$$

and

$$det(\nabla_x Y(x,0))\rho_0(Y(x,0)) = \rho_1(x), \tag{10}$$

thanks to Eq. (7). On the other hand, mass conservation can be written also

$$\partial_t \rho + \nabla_x\cdot(\rho U) = 0, \tag{11}$$

with initial condition $\rho(x,0) = \rho_1(x)$.

The derivative of the cost functional with respect to t is then

$$\begin{aligned}\partial_t I &= \frac{d}{dt}\int_{\Omega_0}\rho_0(\xi)\Big[\|\mathcal{X}(\xi,t)\|^2 - 2\mathcal{X}(\xi,t)\cdot\xi\Big]d\xi = \frac{d}{dt}\int_{\Omega_1}\rho_0(Y(x,t))\|x\|^2 det(\nabla_x Y(x,t))dx - 2\int_{\Omega_0}\rho_0(\xi)\partial_t\mathcal{X}(\xi,t)\cdot\xi d\xi \\ &= \frac{d}{dt}\int_{\Omega_1}\rho(x,t)\|x\|^2 dx - 2\int_{\Omega_1}\rho(x,t)U(x,t)\cdot Y(x,t)dx.\end{aligned} \tag{12}$$

The vector field $Y(x,t)$ can be decomposed as the sum of a divergence-free vector field $Y_\omega(x,t)$ and the gradient of a scalar potential $\Psi(x,t)$, according to the classical Helmoltz decomposition. Hence, $\Psi = \Delta_x^{-1}(\nabla_x \cdot Y)$, $Y_\omega = Y - \nabla_x \Delta_x^{-1}(\nabla_x \cdot Y)$ and $Y_\omega \cdot n = 0$ on $\partial\Omega_1$. We now take

$$U = \frac{Y_\omega}{\rho}, \tag{13}$$

and as a consequence it follows that $\partial_t \rho = 0$, i.e., $\rho(x,t) = \rho_1(x)$. Then, in Eq. (12)

$$\frac{d}{dt}\int_{\Omega_1} \rho(x,t)\|x\|^2 dx = \frac{d}{dt}\int_{\Omega_1} \rho_1(x)\|x\|^2 dx = 0, \tag{14}$$

and

$$\int_{\Omega_1} \rho(x,t)U(x,t)\cdot Y(x,t)dx = \int_{\Omega_1} Y_\omega(x,t)\cdot Y_\omega(x,t)dx + \int_{\partial\Omega_1} \Psi(x,t)Y_\omega(x,t)\cdot n dx = \int_{\Omega_1} Y_\omega(x,t)\cdot Y_\omega(x,t)dx. \tag{15}$$

In summary, by taking $U(x,t)$ as in Eq. (13) we have found an optimal descent direction for the minimization of I:

$$\partial_t I = -2\int_{\Omega_1} Y_\omega(x,t)\cdot Y_\omega(x,t)dx, \tag{16}$$

together with an evolution equation for $Y(x,t)$

$$\partial_t Y + \frac{Y_\omega}{\rho_1}\cdot\nabla_x Y = 0, \tag{17}$$

such that the forward map $\mathcal{X}(\xi,t)$ satisfies

$$det(\nabla_\xi \mathcal{X}(\xi,t))\rho_1(\mathcal{X}(\xi,t)) = \rho_0(\xi), \tag{18}$$

$\forall t \in \mathbb{R}^+$. The minimum of the functional will be reached when $Y_\omega(x,t) = 0$, i.e., when $Y = \nabla_x \Psi$.

This approach to the solution of the optimal mass transfer problem is important because of its links with polar factorization of vector fields [17] and generalized Boussinesq equations [3]. However, from the numerical point of view it suffers from some drawbacks in that one should build $X_t(\xi)$ and this is not always an easy task. Moreover even when Eq. (8) is carefully integrated in time by a high order scheme, mass is not exactly conserved at the discrete level. Therefore if the initial map is far from the minimum and many gradient step iterations are needed, the error in mass conservation may be large. More recently, in [19], it was proposed to directly solve the minimization of (6) under mass conservation constraint, by a sequential quadratic programming approach. This method, however, leads to an optimization problem of the size of the spatial grid resolution.

3. Action minimization

In [20] it is shown that the optimal mass transfer problem is equivalent to the flow of a pressureless ideal compressible fluid. Consider a time-dependent density function $\rho(x,\tau)$ defined in $\mathbb{R}^d$ such that

$$\rho(x,0) = \rho_0(x) \tag{19}$$

and

$$\rho(x,T) = \rho_1(x). \tag{20}$$

The variable τ stands now for time and it plays a different role compared to the parameter t of the previous section. It can be shown (see [20]) that the optimal mass transfer problem is equivalent to the minimization with respect to $U(x,\tau)$ of the time integral of the kinetic energy (the action) associated to the transport:

$$J = \frac{1}{2}\int_0^T \int_{R^d} \rho(x,\tau)\|U(x,\tau)\|^2 dx d\tau, \tag{21}$$

subject to Eqs. (11), (19) and (20).

Introducing a space–time lagrange multiplier $\psi(x,\tau)$, the Euler–Lagrange equations for the constrained minimum of J are

$$\partial_\tau \psi + U\cdot\nabla\psi = \frac{\|U\|^2}{2}, \tag{22}$$

$$U = \nabla\psi. \tag{23}$$

and because initial and final conditions are given for $\rho(x,\tau)$, no conditions are imposed on $\psi(x,\tau)$.

In [16], the action minimization problem under constraint is solved by the Uzawa algorithm. The main disadvantage of this approach is that the discretization of the additional time dimension is such that the size of the discrete problem is multiplied by N, if N is the size of the resolution in one space direction.

An important property of the optimal transport easily follows from this formulation. Indeed, an evolution equation for the potential can be obtained substituting Eq. (23) into Eq. (22):

$$\partial_\tau \psi + \frac{|\nabla \psi|^2}{2} = 0, \tag{24}$$

which is an Hamilton–Jacobi equation that describes a transport along straight lines. This can be seen by taking the gradient of the equation above to obtain

$$\partial_\tau U + (U \cdot \nabla) U = 0, \tag{25}$$

which shows that the velocity $U(x,\tau)$ is constant along a characteristic, i.e., the velocity is constant along rays in space and time. In fact, this means that if ξ is the lagrangian coordinate and $\mathcal{X}(\xi,\tau)$ the map between x and ξ at time τ, we have

$$U(\mathcal{X}(\xi,\tau),\tau) = V(\xi), \tag{26}$$

where $V(\xi)$ is the initial velocity. Deriving the above equation with respect to τ, we find Eq. (25).

4. Mass transport along straight lines

The plan is now to use a lagrangian representation of the density distribution to impose mass conservation. We consider a set of particles such that

$$\rho(x,\tau) \approx \sum_{j=1}^{N_p} c_j(t)\sigma(x - X_j(\tau)) \tag{27}$$

where N_p is the number of particles, X_j is the particle coordinate and $\sigma_j = \sigma(x - X_j(\tau))$ is a regularization of a Dirac mass satisfying

$$\int_{\Omega_r} \sigma(\xi) d\xi = 1, \tag{28}$$

where $\Omega_r \subset \mathbb{R}^d$ is the support of the regularizing kernel σ. Let $\Omega(\tau) = \bigcup_{j=1}^{N_p} \Omega_j(\tau)$, with $\Omega_j(\tau)$ the support of σ_j. We have that

$$\int_{\Omega(\tau)} (\partial_\tau \rho + \nabla \cdot (\rho \nabla \psi)) dx = \frac{d}{d\tau} \int_{\Omega(\tau)} \rho dx, \tag{29}$$

and substituting Eq. (27)

$$\frac{d}{d\tau} \int_{\Omega(\tau)} \rho dx = \frac{d}{d\tau} \int_{\Omega(\tau)} \sum_{j=1}^{N_p} c_j(\tau)\sigma_j dx = \sum_{j=1}^{N_p} \frac{d}{d\tau} \left(c_j(\tau) \int_{\Omega(\tau)} \sigma_j \right) dx, \tag{30}$$

which reduces to

$$\frac{d}{d\tau} \int_{\Omega(\tau)} \rho dx = \sum_{j=1}^{N_p} \partial_\tau c_j(\tau). \tag{31}$$

In the following we assume that

$$\partial_\tau c_j(\tau) = 0, \tag{32}$$

so that the mass conservation equation is identically satisfied. As a consequence, the time invariant quantity c_j can be interpreted as the mass of the jth particle.

Let us now take

$$X_j(\tau) = \xi_j + V(\xi_j)\tau, \tag{33}$$

where ξ_j is the position of the particle at $\tau = 0$ and $V(\xi_j)$ is the initial velocity of the particle. This equation translates the fact that the velocity is constant along straight lines. Thanks to this assumption, Eq. (25) is identically satisfied.

4.1. Reconstruction of the initial condition

The discretization of $\rho(x,\tau)$ verifies the mass constraint by Eq. (32) and the particle trajectories $X_j(\tau)$ are such that $U(x,\tau)$ respects Eq. (25). We now choose the time-invariant coefficients c_j in such a way that the initial mass distribution $\rho(x,0) = \rho_0(x)$ is approximated in a least-square sense.

Given $\rho(x,0)$ on a regular cartesian mesh, a simple choice consists in placing the particles in the nodes where $\rho(x,0) > \delta_0$, δ_0 being the smallest density that is considered. At the discrete level, the approximation problem is then formulated as an optimization problem for the c_j:

$$c_j = \arg\left\{ \min_{d_j} \sum_{k=1}^{N_g} \left[\rho(x_k, 0) - \sum_{j=1}^{N_p} d_j \sigma(x_k - X_j(0)) \right]^2 \right\}, \tag{34}$$

where N_g is the number of grid points where the error is evaluated. The particle initial positions $X_j(0) = \xi_j$ are given and coincide with the grid points where the density is above a certain threshold. Other more sophisticated choices, like for example adapting the particle distribution to $\rho(x,0)$, can lead to better accuracy. However, if the initial position of the particles is given, the reconstruction of the initial density distribution always amounts to a quadratic optimization problem in the coefficients c_j that can be solved by a linear system. The computational cost of this step is negligible since the size of the problem is $N_p \times N_p$ and the number of particles is usually of the order of 10^3–10^6, according to resolution and the number of space dimensions.

4.2. Potential velocity field and reconstruction of the final condition

The mass c_j of each particle is now determined from the approximation of the initial condition. The particles move along straight lines and the particle mass remains constant along these trajectories. Two conditions for the minimum of (21) are hence satisfied. We still have to enforce that the velocity field is potential and that the final condition on the density distribution is verified. In order to do so, we assume that the components of the velocity are expressed as centered finite differences in the respective directions of a scalar function whose values on the grid are ψ_l. Denoting by D_{jl} the elements of the discrete centered gradient operator, we have that the velocity of each particle is $V_j = \sum_{l=1}^{N_d} D_{jl}\psi_l$, where N_d is of the order of N_p. Next, an optimization problem with respect to ψ_l is solved to approximate the final mass distribution. We have

$$\psi_l = \arg\left\{ \min_{\Psi_l} \mathcal{E}(\Psi_l) \right\} = \arg\left\{ \min_{\Psi_l} \sum_{k=1}^{N_g} \left[\rho(x_k, T) - \sum_{j=1}^{N_p} c_j \sigma(x_k - \xi_j - \sum_{l=1}^{N_d} D_{jl}\Psi_l T) \right]^2 \right\}. \tag{35}$$

The gradient of the above function can easily be computed so that the numerical solution of this problem is solved by a steepest descent method or by quasi-Newton iterations.

Mollifying kernels $\sigma(\xi)$ with compact support can be used, although the support must be large enough to keep some desirable properties. In cases in which, for example, a fragmentation process takes place, or simply in cases in which the density supports at times $\tau = 0$ and $\tau = T$ have null intersection, compact support kernels with small enough support may result in having $\partial\mathcal{E}/\partial\Psi_l = 0$ from the first optimization step because the error vanishes where the support of the kernel is non-zero and vice versa.

In order to possibly speed up convergence toward the minimum, a penalization can be added to $\mathcal{E}(\Psi_l)$:

$$\mathcal{E}_p(\Psi_l) = \mathcal{E}(\Psi_l) + \beta \sum_{j}^{N_p} c_j \frac{\|\sum_{l=1}^{N_d} D_{jl}\Psi_l\|^2}{2}, \tag{36}$$

where $\beta \in \mathbb{R}^+$. The actual effect on convergence of the penalization term is studied in the numerical experiments hereafter.

In summary, in the discrete problem that we have formulated mass conservation, mass transport along straight lines and potential velocity field are exactly satisfied at the discrete level. The initial and final mass distribution are approximated in a least square sense. Of course, the problem solution is independent of T. We kept the notion of the time variable τ in case the intermediate states of the mapping are to be computed. This is an inexpensive task thanks to Eq. (33).

5. Preliminary numerical tests

In the following numerical experiments we use the gaussian kernel

$$\sigma(\xi) = \frac{1}{(\ell\sqrt{\pi})^d} \exp\left(-\frac{|\xi|^2}{\ell^2} \right), \tag{37}$$

where ℓ is the kernel characteristic length.

The first test is relative to a problem where the initial density distribution is uniform and the final one shows concentric compressions and rarefactions. The error in the initial and final density distributions is studied as a function of the discretization parameters. The other examples are aimed at analyzing the performance of the proposed solution method in more critical cases where density distributions with sharp boundaries are given or mass fragmentation phenomena occur. These cases are paradigmatic of situations that are encountered in medical imagery problems. The last two applications show results in this sense.

In all test cases the reconstruction of the final condition was attained by a constant step gradient descent method. The most computationally intense simulation, the two-dimensional mass splitting phenomenon, takes about 1 h on a standard laptop. In the next pictures, when isocontours are shown, 25 levels equally distributed between the minimum and the maximum value of the scale are drawn.

5.1. Recovering a wavy density distribution

The initial density distribution is $\rho(x,0) = 1$ for $x \in [0,1] \times [0,1]$. We consider a velocity distribution

$$U(x,0) = K\nabla(\sin(2\pi\omega\|x\|^2)\exp(-(\|x\|^2/\eta^2))),$$

where $K = 1.00e-2$, $\omega = 4$ and $\eta = 0.25$. The initial density distribution and the velocity field are propagated in time according to Eqs. (11) and (25). Once the density distribution at time $T = 1$ is found, we solve the optimal mass transfer problem. In Fig. 1 the initial and final density distributions are represented. Rarefactions and compression zones can be identified according to the grey scale. The space resolution of the figure is 200×100. In this test case the resolution is intentionally not uniform.

Three different grid resolutions were considered, 100×50, 200×100, and 400×200. For such resolutions, placing particles where density is larger than the density maximum divided by 1000, leads to the following number of particles: 1.25e3, 5.00e3 and 2.00e4, respectively. The minimization of $\mathcal{E}$ was stopped when the difference between the gradient norm of two subsequent iterations was less than 1.0e−4. The results are presented in Tables 1–3.

First of all we remark that β has a reduced influence on the results. ϵ_T slightly decreases with β increasing, but then this effect is saturated. Of course ϵ_0 is independent of β. We have that $\epsilon_0 \to 0$ as $\ell \to 0$, since in the limit case c_j equals the local value of the initial density distribution. Increasing the grid resolution, the number of particles increases and the error on the final resolution is systematically decreasing.

The error ϵ_T is not monotonically decreasing with the kernel length. There is a trade off between accuracy on the initial and final density distributions, depending on the ratio between the grid size and the kernel length. However, the optimal kernel length decreases with increasing grid resolution, as it should for consistency. No significant trend appears in the number of iterations except that in general the number of iterations increases as β does.

The values of $|\rho(x_k,T) - \sum_j^{N_p} c_j\sigma(x_k - \xi_j - V_jT)|/\rho(x_k,T)$ are shown in Fig. 2, when the resolutions used are 200×100 and 400×200 and $\ell = 1.00e-2$, $\beta = 5.0e-3$. The error is concentrated on the boundaries and in the higher density zones.

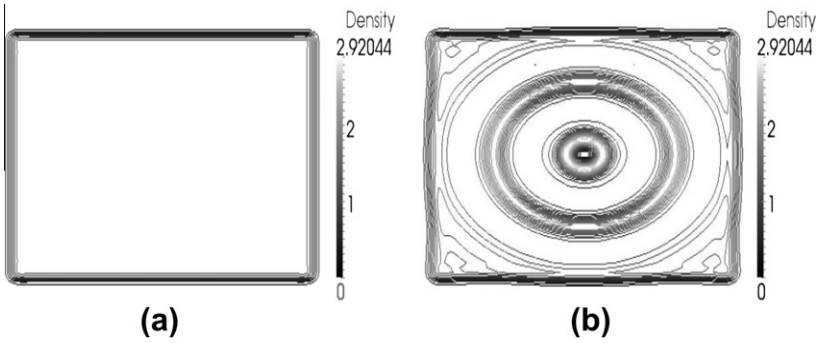

Fig. 1. Density distribution at time (a) $T = 0$, (b) $T = 1$.

Table 1
Resolution 100×50. ϵ_0 is the L^2 relative error on the initial density distribution, ϵ_T on the final. N_i is the number of iterations to minimize $\mathcal{E}$, i.e., such that the gradient norm variation is below 1.0e−4.

ℓ	β	ϵ_0	ϵ_T	N_i
2.00e−2	5.0e−4	6.776e−4	6.078e−4	230
1.80e−2	5.0e−4	6.120e−4	4.437e−4	269
1.60e−2	5.0e−4	5.398e−4	3.185e−4	336
1.40e−2	5.0e−4	4.648e−4	2.821e−4	409
1.20e−2	5.0e−4	3.871e−4	4.076e−4	522
2.00e−2	1.0e−3	6.776e−4	5.882e−4	328
1.80e−2	1.0e−3	6.120e−4	4.078e−4	380
1.60e−2	1.0e−3	5.398e−4	2.641e−4	627
1.40e−2	1.0e−3	4.648e−4	2.326e−4	508
1.20e−2	1.0e−3	3.871e−4	3.748e−4	623
2.00e−2	5.0e−3	6.776e−4	5.434e−4	561
1.80e−2	5.0e−3	6.120e−4	4.040e−4	481
1.60e−2	5.0e−3	5.398e−4	2.622e−4	572
1.40e−2	5.0e−3	4.648e−4	2.302e−4	493
1.20e−2	5.0e−3	3.871e−4	3.727e−4	620

Table 2
Resolution 200×100.

ℓ	β	ϵ_0	ϵ_T	N_i
1.40e−2	5.0e−4	3.979e−4	2.633e−4	173
1.20e−2	5.0e−4	3.729e−4	1.675e−4	177
1.00e−2	5.0e−4	3.249e−4	8.646e−5	245
0.80e−2	5.0e−4	2.566e−4	5.408e−5	399
0.60e−2	5.0e−4	1.873e−4	4.704e−4	140
1.40e−2	1.0e−3	3.979e−4	2.655e−4	161
1.20e−2	1.0e−3	3.729e−4	1.678e−4	184
1.00e−2	1.0e−3	3.249e−4	8.852e−5	236
0.80e−2	1.0e−3	2.566e−4	5.703e−5	393
0.60e−2	1.0e−3	1.873e−4	4.859e−4	142
1.40e−2	5.0e−3	3.979e−4	2.673e−4	189
1.20e−2	5.0e−3	3.729e−4	1.768e−4	196
1.00e−2	5.0e−3	3.249e−4	1.019e−5	244
0.80e−2	5.0e−3	2.566e−4	7.522e−5	360
0.60e−2	5.0e−3	1.873e−4	3.855e−4	230

Table 3
Resolution 400×200.

ℓ	β	ϵ_0	ϵ_T	N_i
1.00e−2	5.0e−4	8.288e−5	3.737e−5	230
0.85e−2	5.0e−4	7.108e−5	2.754e−5	204
0.70e−2	5.0e−4	5.921e−5	3.420e−5	307
0.55e−2	5.0e−4	4.790e−5	9.985e−5	390
0.40e−2	5.0e−4	3.573e−5	2.704e−4	193
1.00e−2	1.0e−3	8.288e−5	3.821e−5	219
0.85e−2	1.0e−3	7.108e−5	2.770e−5	213
0.70e−2	1.0e−3	5.921e−5	3.517e−5	326
0.55e−2	1.0e−3	4.790e−5	9.998e−5	410
0.40e−2	1.0e−3	3.573e−5	3.004e−4	195
1.00e−2	5.0e−3	8.288e−5	3.975e−5	250
0.85e−2	5.0e−3	7.108e−5	2.821e−5	245
0.70e−2	5.0e−3	5.921e−5	3.519e−5	278
0.55e−2	5.0e−3	4.790e−5	1.021e−4	334
0.40e−2	5.0e−3	3.573e−5	4.455e−4	227

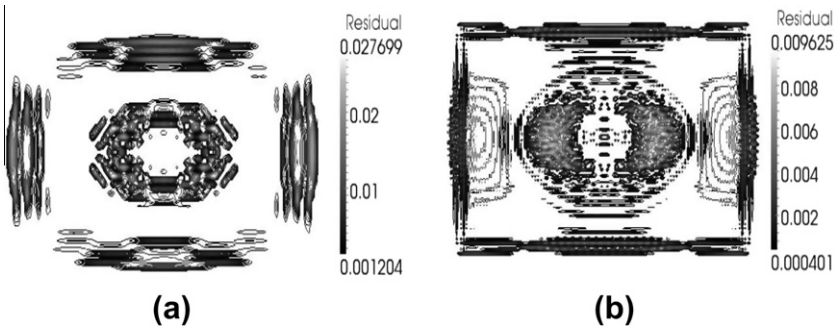

Fig. 2. Relative error on the final density distribution. Resolution: (a) 200×100, (b) 400×200.

5.2. Mass splitting: recovering a one-dimensional exact map

In this subsection we study the accuracy of the scheme proposed in a case where the exact map $X^*(\xi)$ is known. The difficulty is that the mass splits and that the initial and final densities are not uniformly differentiable, whereas the kernels used to represent the density distributions are uniformly differentiable. The initial density, of mass 1, is a hat function defined between -1 and 1. The final density distribution, of equal mass, is represented by two hat functions, symmetrically placed

about the vertical axis and defined between -1 and 0, 0 and 1, respectively. The exact optimal map can be determined in this case by simply integrating the jacobian equation.

In Fig. 3 the results obtained with 80 grid points are contrasted to the exact solution. On the left, the initial density distribution, on the right the final one. In Table 4 the results of the numerical solution are presented. In particular, we consider the error ϵ_T defined as before. Also, we consider the error ϵ_a, the L^2 norm of the difference between the exact solution and the result of the simulation, computed with an higher order quadrature. In this case it is possible to compute such an error since both the exact solution and the one resulting from the simulation are defined for all values between -1 and 1. Finally, we can compute E_{map}, the L^2 error between the exact map and the one obtained in the simulation. Since we actually compute the potential, this is a more stringent error since it is relative to a differentiated quantity. As before, the errors systematically decrease with increasing grid resolution. Also, the optimal kernel length decreases as the number of points increases.

5.3. Two-dimensional mass splitting

We investigate the two dimensional case of a final density distribution obtained using a potential that is just C^0. The velocity field is discontinuous and hence the mass distribution is fragmented. In Fig. 4 the density at time $T = 0$ and $T = 1$ are shown. At the beginning, the distribution is a normalized paraboloid. The potential from which the velocity field derives is $0.01|x - 0.1\sin(4\pi y) - 0.5|$.

The image is assigned on a 200×100 grid and the number of particles used is about 5e3. The number of iterations for the algorithm to converge is approximately 1500. In Fig. 5(a) the solution obtained by numerically solving the optimal transport problem is shown. All the details of the fragmentation process are correctly recovered. In Fig. 5(b) the relative error on the final image in the case of 200×100 resolution is represented. The relative error is not concentrated along the singularity, but on the external boundary, since there the density values are close to 0.

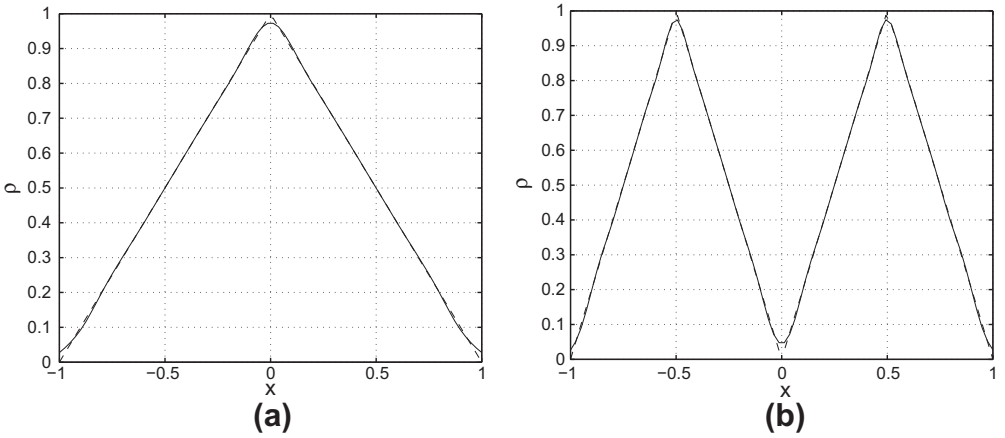

Fig. 3. Density distribution at time (a) $T = 0$, (b) $T = 1$. The grid resolution is $N = 80$.

Table 4
N_p is the number of particles, ℓ is the kernel length, ϵ_T is the error computed on the actual grid, ϵ_a is the error computed with respect to the exact solution, and E_p is the error on the mapping, i.e., on the gradient of the exact potential. The regularization was set to $\beta = 5e{-}4$.

N_p	ℓ	ϵ_T	ϵ_a	E_{map}
10	2.00e−1	1.74e−2	8.30e−2	1.16e−1
10	1.75e−1	1.63e−2	8.44e−2	1.14e−1
10	1.60e−1	1.71e−2	8.51e−2	1.15e−1
20	1.60e−1	1.52e−2	3.39e−2	0.92e−1
20	1.50e−1	1.43e−2	3.10e−2	0.91e−1
20	1.40e−1	1.21e−2	2.93e−2	0.90e−1
40	1.15e−1	1.00e−2	1.60e−2	0.79e−1
40	1.10e−1	0.95e−2	1.51e−2	0.77e−1
40	1.05e−1	0.93e−2	1.46e−2	0.74e−1
80	8.50e−2	8.91e−3	9.98e−3	0.55e−1
80	7.50e−2	7.13e−3	8.41e−3	0.51e−1
80	7.00e−2	6.75e−3	7.54e−3	0.53e−1

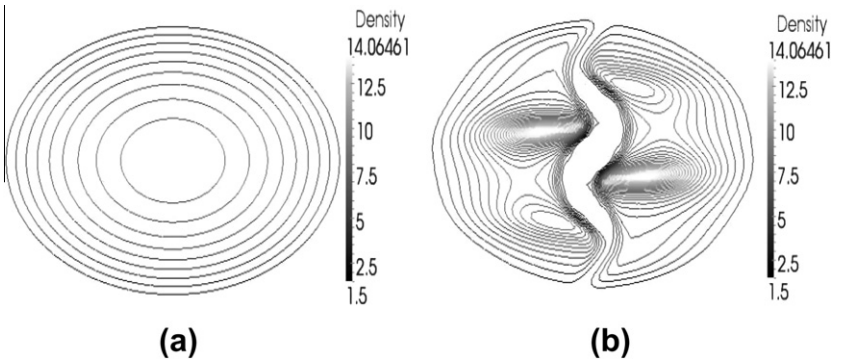

Fig. 4. Reference density distributions, resolution 200×100, at time (a) $T = 0$, (b) $T = 1$.

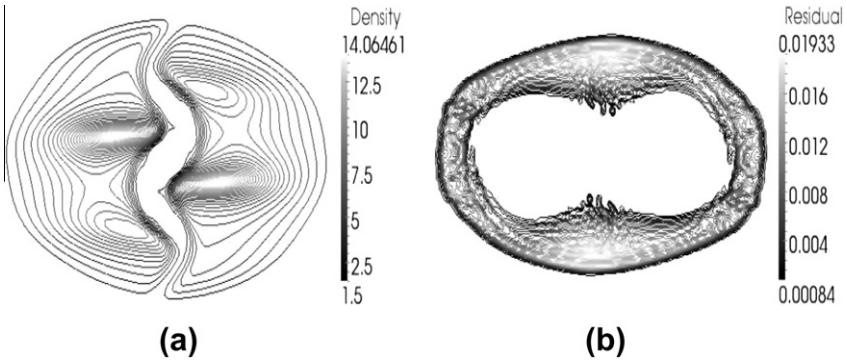

Fig. 5. (a) Density at time $T = 1$ for the computed optimal solution, (b) relative error on the final image.

5.4. Application to medical imagery

One possible application of optimal transport techniques concerns medical imagery and, in particular, non-rigid registration, see for example [14]. Image registration is the process of establishing a common geometric reference frame between two or more data sets possibly taken at different times.

Registration has a substantial recent literature devoted to it, with numerous approaches effective in varying situations, as described in [21]. These range from optical flow to computational fluid dynamics, to various types of warping methodologies. One class of methods is based on variational techniques, where the characterization of the desired transformation is embodied in the definition of the functional to be minimized. A mass preserving mapping that minimizes the distance may be of practical interest thanks to certain desirable properties: it is parameter free, symmetrical and the minimizer of the distance functional involved is unique. However, in the present context, image registration is used moreover as a challenging numerical illustration.

The first example discussed is relative to a thorax scan presenting lung noduli, while the second example is relative to the morphing of an image showing lung tissue attacked by cancer.

5.4.1. Thorax scan

In Fig. 6 the initial and final thorax scans are shown. The objective is to find an optimal mapping between the two geometries. The images have a resolution of 200×200 and the number of particles used is approximately 2.0e4. The number of iterations to reach convergence, due to the complexity of the geometry was 2500.

In Fig. 7 the result of the computation is presented. On the left the solution obtained by applying the transformation to the first image is shown, on the right there is the actual scan. The main difference is in the normalization of grey scale: in order to perform a registration based on the optimal transport problem, the mass of the two density distributions must be the same and therefore the grey scale of the final image is modified so that it has the same "mass" of the initial image. Despite the complexity of the geometry, the agreement is quite good.

In Fig. 8(a) the relative error on the final image is shown. It is higher in the region where there are sharp boundaries, as expected. The error is basically due to the gaussian kernel spreading sharp boundaries. In Fig. 8(b) the potential of the transformation is shown.

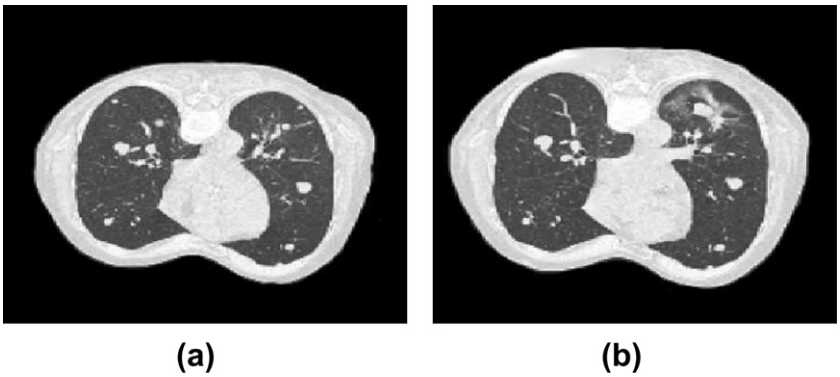

Fig. 6. Thorax image scan at (a) $T = 0$, (b) $T = 1$ (corresponding to an evolution of six months); Courtesy Institute Bergonié, Bordeaux.

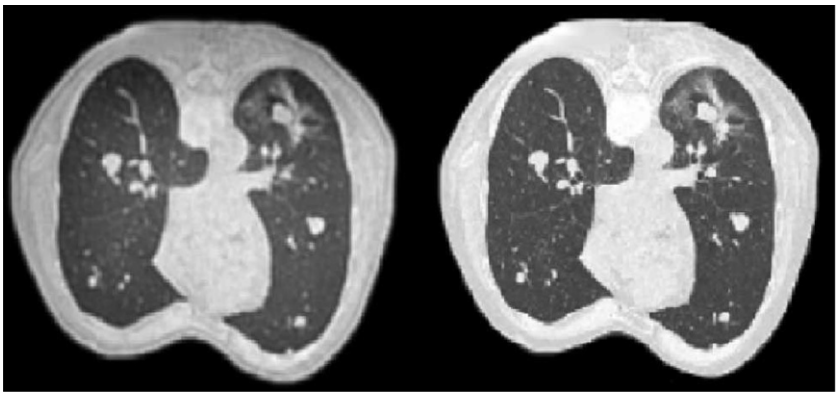

Fig. 7. On the left: optimal transport result at $T = 1$; on the right: the true image at the same time.

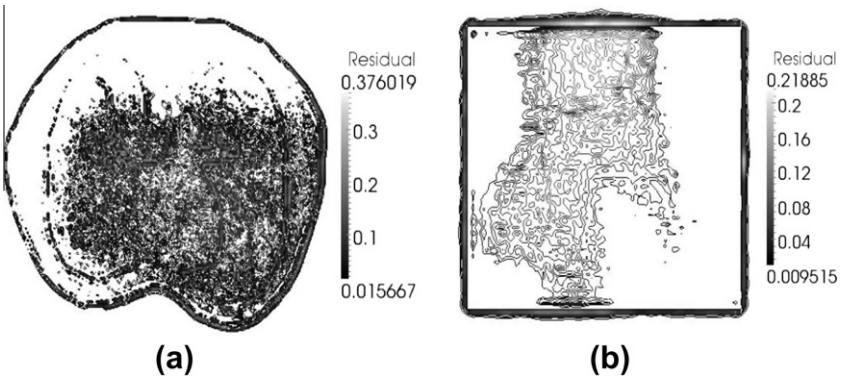

Fig. 8. Residual of the final image in the case of (a) thorax non-rigid registration, (b) tissue mapping for a regressing tumor.

5.4.2. Tissue morphing

A growing tumor can be associated to a certain displacement field. This field can be computed as the optimal plan that realizes the mapping between the scans. A portion of a lung scan representing tissue affected by cancer is considered and in Fig. 9 two images are shown. The first image represents the original tumor ($T = 0$), which is then treated and its area decreases ($T = 1$). The space resolution of these images is 128×128 pixels, which is the original resolution of the scan. In Fig. 10 the result of the optimal flow is compared to the actual image. The grey scale has slightly changed and this is due to the fact that the mass (the integral of the density over the domain) is decreased when the tumor is collapsed, so that a renormalization of the grey scale was necessary in order to enforce the mass conservation constraint. Apart from this, the agreement is good and no particular error structure emerges from the computation.

In Fig. 8(b) the relative error with respect to the normalized image is shown.

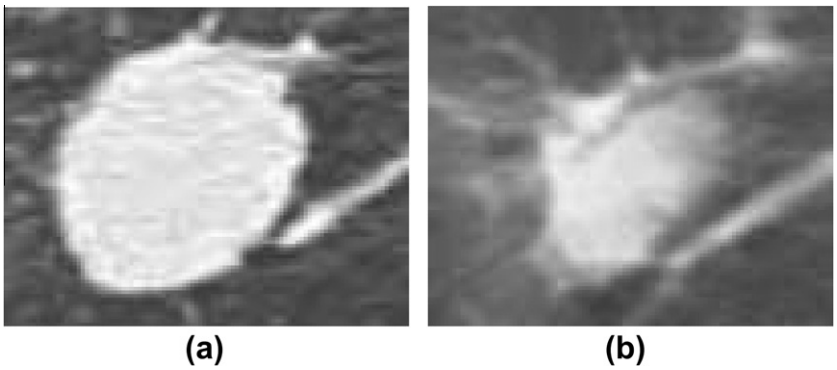

Fig. 9. Tissue scan at (a) $T = 0$, (b) $T = 1$ (corresponding to an evolution of three months); Courtesy Institute Bergonié, Bordeaux.

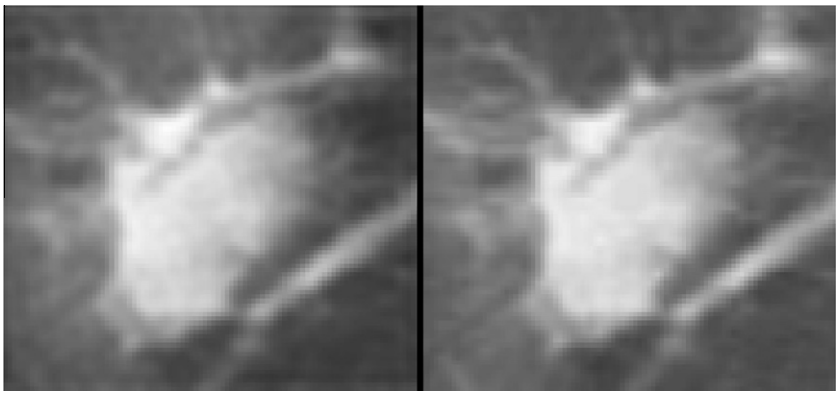

Fig. 10. On the left: Monge result at $T = 1$; on the right: the true image at the same time.

6. A three-dimensional application

The objective of this section is to provide a discussion of the present approach computational viability to recover a non-trivial density distribution. Again, we take an example from medical imagery. The problem is to map a uniform mass distribution in a cube to a density distribution that corresponds to the magnetic resonance imaging (MRI) of a human head.

Gaussian kernels evaluated on a sub-domain of 512 points and three different grid resolutions were employed: 15^3, 30^3 and 60^3. The simulations were stopped when the L^2 norm of the residual, as defined for the previous numerical tests, was divided by a factor 100 with respect to its initial value.

In Fig. 11 the initial and final densities are represented at the highest resolution and in Fig. 12 cuts in different planes of the same distribution show the complexity of the geometry to be recovered. In Fig. 13 planes with the isocontours of the residual are represented, superposed to the solution. The residual is negligible everywhere except for some spots, corresponding to the sharpest details of the target distribution.

In Table 5 we report the computational time per gradient iteration (T_{it}), and the number of gradient iterations (It). The computations were performed on a standard laptop computer. As before, N_p is the number of particles, ℓ the kernel length and β, the penalisation, is equal to 2e−2 for all the resolutions. If the number of particles is N_p and the grid points are N, using global kernels (kernels whose support is the whole domain) leads to a computational time per gradient iteration scaling with $N_p \times N \approx N^2$. But, as discussed, a simple truncated kernel evaluation leads to a linear scaling with N_p. This trend is indeed confirmed by the evolution of the computational time per iteration as a function of the number of particles in Table 5. The AHT method [15] has also a computational cost per iteration which is proportional to N, whereas in the Uzawa method [16] the cost of one iteration scales more than linearly because the computation of the gradient is obtained thanks to a fast Poisson solver in $d + 1$ dimensions.

Overall, the possible advantages of the present method reside mainly in the exact mass conservation and in the fact that N_p might be significantly smaller than N, since a particle is present only where mass is present. The actual computational bottle neck is rather linked to the number of gradient iterations to get to the minimum. However, this is a common issue to all methods based on optimization, i.e., the AHT and the Uzawa scheme. Of course, a full discussion of this issue should involve the initial image, i.e., how far from the solution the minimization starts, the width of the scale spectrum of the target

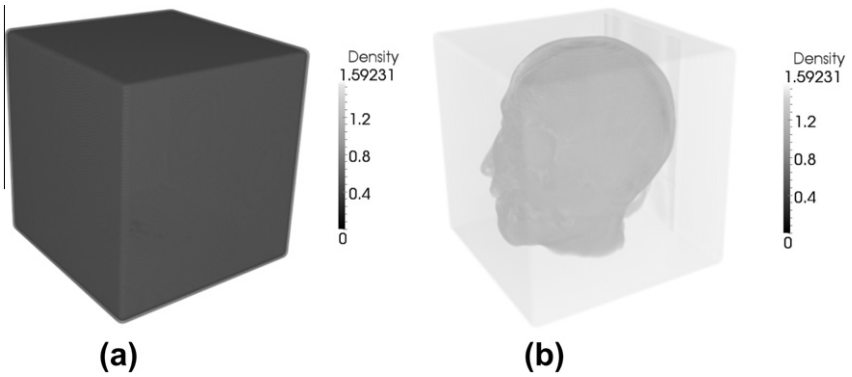

Fig. 11. Density distribution at time (a) $T = 0$, (b) $T = 1$.

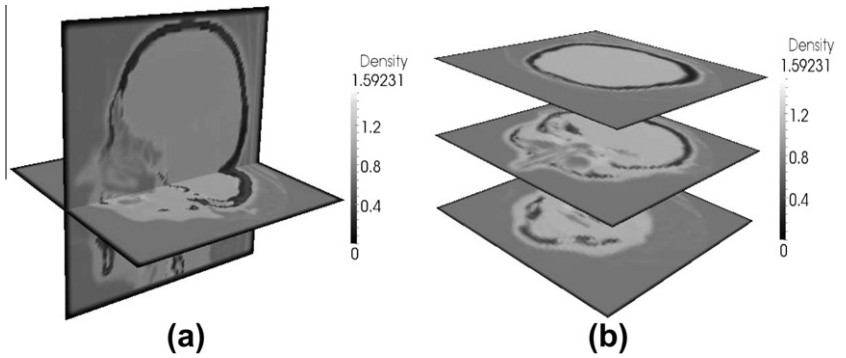

Fig. 12. Slices of the final density distribution: (a) *XZ* planes, (b) *Z* planes.

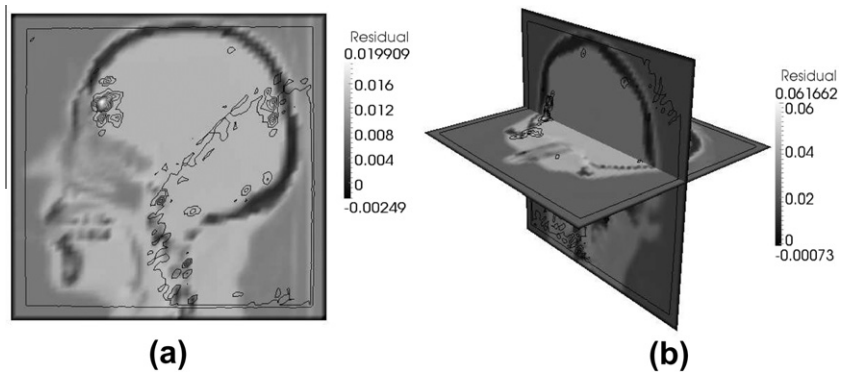

Fig. 13. (a) Contour lines of the residual distribution in the *X* plane; (b) contours of residual in *YZ* planes.

Table 5
Computational cost per gradient iteration (T_{it}), and the number of gradient iterations (It).

N_p	ℓ	T_{it}	It
15^3	1.2e−3	1.2s	1000
30^3	2.4e−4	9s	1100
60^3	1.6e−4	72s	4000

image, and the actual minimization method employed. In this direction, a promising approach seems to be a multilevel optimization where the initial density distribution on a finer grid is obtained from the solution of a Monge problem on a coarser grid.

7. Conclusions

We described a lagrangian method to solve the L^2 optimal transport problem. The main idea is to identically satisfy the mass conservation equation as well as the Hamilton–Jacobi equation for the potential. The initial and final conditions on the density distribution are relaxed, and they are satisfied in a least-square sense. This approach does not require the numerical solution of any PDE, since the problem solution amounts to the minimization of a non-linear function with respect to scalar values (the potential) defined on a subset of the discretization grid. In the numerical tests, we show that the error in the potential or in the representation of the initial and final density distribution systematically decreases by increasing the number of discretization points, and hence, of mass particles. The test cases show also that the method recovers critical mass splitting phenomena, even with sharp boundaries. Some paradigmatic applications in medical imagery are also presented in order to investigate performance in real life problems. One draw back of this approach is that, because of the regularization kernel, sharp images may be blured if the resolution is not high enough. However, the examples presented suggest that such phenomenon is acceptable, and since the kernel characteristic length can be decreased as the grid refinement is increased, this effect is less and less relevant. Finally, the formulation described here is one of the possible implementations of the idea of letting mass particles move along rays to satisfy mass conservation and the non-linear transport equations resulting from the minimum conditions. Other kernels, or other minimization problems taking the initial and final conditions as hard constraints can be defined based on the same approximation. These approaches are currently under scrutiny.

References

[1] G. Monge, Memoire sur la theorie des deblais et des remblais, Histoire de l'Academie des Sciences de Paris, 1781.
[2] L.V. Kantorovich, On the transfer of masses, Doklady Akademii Nauk SSSR 37 (1942) 227–229.
[3] Y. Brenier, Optimal transport, convection, magnetic relaxation and generalized Boussinesq equations, Journal of Nonlinear Science 19 (2009) 547–570.
[4] G. Crasta, A. Malusa, On a system of partial differential equations of Monge–Kantorovich type, Journal of Differential Equations 235 (2007) 484–509.
[5] G. Crasta, A. Malusa, A variational approach to the macroscopic electrodynamics of anisotropic hard superconductors, Archive for Rational Mechanics and Analysis 192 (2009) 87–115.
[6] L.C. Evans, M. Feldman, Fast/slow diffusion and collapsing sandpiles, Journal of Differential Equations 137 (1997) 166–209.
[7] B. Maury, J. Venel, A mathematical framework for a crowd motion model, Comptes Rendu Mathematiques 346 (2008) 1245–1250.
[8] C. Villani, Topics in Optimal Transportation, American Mathematical Society, 2003.
[9] C. Villani, Optimal Transport, Old and New, Springer, Verlag, 2009.
[10] L.C. Evans. Partial differential equations and Monge–Kantorovich mass transfer. Technical report, University of California, Berkley, 1997.
[11] M.J.P. Cullen, R.J. Purser, A duality principle in semi-geostrophic theory, Journal of Atmospheric Sciences 44 (1987) 3449–3468.
[12] W. Gangbo, V. Oliker, Existence of optimal maps in the reflector-type problems, ESAIM COCV 13 (2007) 93–106.
[13] P. Olver, A. Tannenbaum, Mathematical Methods in Computer Vision, Springer-Verlag, 2004.
[14] T. Rehman, E. Haber, G. Pryor, J. Melonakos, A. Tannenbaum, 3D nonrigid registration via optimal mass transport on the GPU, Medical Image Analysis 13 (2009) 931–940.
[15] S. Angenent, S. Haker, A. Tennenbaum, Minimizing flows for the Monge–Kantorovich problem, SIAM Journal Mathematical Analysis 35 (2003) 61–97.
[16] J-D. Benamou, Y. Brenier, K. Guittet, The Monge–Kantorovitch mass transfer and its computational fluid mechanics formulation, International Journal for Numerical Methods in Fluids 40 (2002) 21–30.
[17] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Communication in Pure and Applied Mathematics 64 (1991) 375–417.
[18] W. Gangbo, An elementary proof of the polar factorization of vector-valued function, Archive for Rational Mechanics and Analysis 128 (1994) 381–399.
[19] E. Haber, T. Rehmam, A. Tannenbaum, An efficient numerical method for the solution of the L^2 optimal transfer problem, SIAM Journal on Scientific Computing 32 (2010) 197–211.
[20] J-D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, Numerische Matematik 84 (2000) 375–393.
[21] Jan Modersitzki, Numerical Methods for Image Registration, Oxford University Press, 2004.