Primal-dual formulation of the Dynamic Optimal Transport using Helmholtz-Hodge decomposition

Morgane Henry, Emmanuel Maitre, Valérie Perrier

Primal-dual formulation of the Dynamic Optimal Transport using Helmholtz-Hodge decomposition

Morgane Henry*, Emmanuel Maitre*, and Valérie Perrier*

Abstract.
This work deals with the resolution of the dynamic optimal transport (OT) problem between 1D or 2D images in the fluid mechanics framework of Benamou-Brenier [6]. The numerical resolution of this dynamic formulation of OT, despite the successful application of proximal methods [36] is still computationally demanding. This is partly due to a space-time Laplace operator to be solved at each iteration, to project back to a divergence free space. In this paper, we develop a method using the Helmholtz-Hodge decomposition [23] in order to enforce the divergence-free constraint throughout the iterations. We prove that the functional we consider has better convexity properties on the set of constraints. In particular we explain that in 1D+time, this formulation is equivalent to the resolution of a minimal surface equation. We then adapt the first order primal-dual algorithm for convex problems of Chambolle and Pock [12] to solve this new problem, leading to an algorithm easy to implement. Besides, numerical experiments demonstrate that this algorithm is faster than state of the art methods for dynamic optimal transport [36] and efficient with real-sized images.

Key words.
Convex optimization, optimal transport, proximal splitting, image processing, Helmholtz-Hodge decomposition, minimal surface

AMS subject classifications. 65K10, 49M25, 49M29, 76B47

1. Introduction.

Optimal transport is a growing field, having numerous applications in different domains such as economics [13, 22], machine learning [42, 33, 1, 43] or partial differential equations [10, 30, 9, 8]. Above all, a lot of applications have arisen in image processing [32], such as color image processing [20, 21], color transfer [40], segmentation [38] or image interpolation [36, 14, 15]. This last application is particularly relevant for the optimal transport, since it defines a metric between densities [45]. Several recent works investigate new formulations for the optimal transport to make the interpolation more physical [7, 34]. From the numerical point of view, while the application of proximal methods [36] was successful, the development of efficient new algorithms for the calculation of the dynamic optimal transport between two densities, due to the lack of strict convexity, is challenging (see [37] for an up-to-date tour on optimal transport).

In this paper, we focus on the image interpolation problem and the numerical resolution of the L^2-optimal transport problem, for which few numerical methods have been developed so far [6, 24, 36]. The pioneering work of Benamou-Brenier [6] places the problem in the context of fluid mechanics by adding a time dimension, leading to a new formulation called the dynamic optimal transport.

The Benamou and Brenier algorithm is based on the minimization of a functional which preserves the mass, through an augmented Lagrangian approach. In contrast we propose

*Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France (emmanuel.maitre@univ-grenoble-alpes.fr,valerie.perrier@univ-grenoble-alpes.fr).

to work directly in the space of constraints for the functional to minimize. Indeed, existing algorithms [36, 6, 21] require a projection onto the divergence-free constraint at each iteration, which amounts to solving a 3D Poisson equation for 2D images. The fact that the functional has better convexity properties in the set of constraints, which we will justify in this paper, motivates our approach. To work in this space, defined by a divergence-free and regularity constraint on the density and the velocity fields with boundary conditions, we use the Helmholtz-Hodge decomposition: it separates any vector field into a curl-free and a divergence-free component, and is often used in the context of partial differential equations [23, 31]. This change of unknown leads to a new formulation of the problem, which in 1D + time, is equivalent to the resolution of a minimal surface equation on each level set of the potential, equipped with appropriate Dirichlet boundary conditions. Another approach to handle the new formulation is to use the first order primal dual algorithm for convex problems developed by Chambolle and Pock [12], which can be easily adapted in our case. The Chambolle-Pock method is nowadays widely used [16, 26], since it leads to fast implementations and can be speed up on parallel architectures. Therefore our method will provide a fast algorithm, simple to implement on imaging problems.
This paper is organized as follows. In section 2, the dynamic optimal transport framework is introduced and the convexity of the functional in the set of constraints is studied, justifying our approach. Then, introducing the Helmholtz-Hodge decomposition, we reformulate the problem directly in the set of constraints. In 1D+time, we establish that the solution satisfies a minimal surface equation. Section 3 is dedicated to the application of a primal-dual algorithm adapted for our functional. Finally numerical experiments are conducted to compare our algorithm to state of the art on several test cases, including real images, proving the validity and efficiency of our method. Part of this work has been published in a conference paper [[28]. In the present paper, we added theoretical developments on the functional on the constraint space, an equivalent formulation in the 1D+t case in terms of a minimal surface resolution, leading to an alternative algorithm, and several numerical experiments.

2. New formulations of the Monge-Kantorovich problem.

2.1. Introduction. Let $\Omega = (0,1)^n$, and $(\rho_0, \rho_1) \in \left(L^2(\Omega)\right)^2$, with $n \in \mathbb{N}^*$, be two positive, bounded densities with

$$\int_\Omega \rho_0(x)dx = \int_\Omega \rho_1(x)dx = 1.$$

If $|.|$ denotes the Euclidean norm in $\mathbb{R}^n$, the L^2-Wasserstein distance (see for example [45]) between ρ_0 and ρ_1 is defined by

$$d_2(\rho_0,\rho_1)^2 = \inf_M \int_\Omega |M(x) - x|^2 \rho_0(x)dx,$$

where the infimum is taken among the maps M transferring ρ_0 to ρ_1, which means that $\forall A \subset \Omega$, $\int_{x\in A} \rho_1(x)dx = \int_{M(x)\in A} \rho_0(x)dx$. The Monge-Kantorovich problem (MKP) consists in finding an application M (the optimal transport) which realizes this infimum.

Benamou and Brenier [6] placed the problem in the continuum mechanics framework. Let us consider a time interval $(0,1)$ for sake of simplicity, set $Q=(0,1)\times\Omega$ and

$$V(Q)=\{f\in(L^2(Q))^{1+n},\ \mathrm{div}_{t,x} f=0\}. \tag{1}$$

Let us introduce the time-dependent density $\rho(t,x)\geq 0$ and the vector field $v(t,x)\in\mathbb{R}^n$ verifying the continuity equation

$$\mathrm{div}_{t,x}(\rho(t,x),\rho v(t,x))=\partial_t\rho(t,x)+\nabla_x\cdot(\rho v)(t,x)=0 \tag{2}$$

for $t\in(0,1)$ and $x\in\Omega$, equipped with the initial, final and boundary conditions

$$\rho(0,x)=\rho_0(x),\ \rho(1,x)=\rho_1(x),\ \forall x\in\Omega, \tag{3}$$

$$\rho v(t,x)\cdot\nu_\Omega=0,\ \forall t\in(0,1),\ x\in\partial\Omega, \tag{4}$$

where ν_Ω is the outward normal of Ω. As proven in [6] (see also [45]), the square of the L^2-Wasserstein distance between ρ_0 and ρ_1 verifies

$$d_2(\rho_0,\rho_1)^2=\inf\int_0^1\int_\Omega\rho(t,x)|v(t,x)|^2dxdt,$$

where the infimum is taken among all ρ,v satisfying (2) and (3). To obtain a convex problem with linear constraints, Benamou and Brenier introduced the momentum $m=\rho v$ and obtained the following formulation

$$\min_{(\rho,m)\in C}\mathcal{J}(\rho,m)\quad\text{where}\quad\mathcal{J}(\rho,m)=\int_0^1\int_\Omega J(\rho(t,x),m(t,x))dxdt, \tag{5}$$

with

$$\forall(\rho,m)\in\mathbb{R}_+\times\mathbb{R}^n,\ J(\rho,m)=\begin{cases}\frac{|m|^2}{2\rho}, & \text{if } \rho>0,\\ 0, & \text{if } (\rho,m)=(0,0),\\ +\infty, & \text{otherwise },\end{cases} \tag{6}$$

and the set of affine constraints reads

$$C:=\{(\rho,m);\ \mathrm{div}_{t,x}(\rho,m)=0,\ m(.,x).\nu_\Omega=0,\ \forall x\in\partial\Omega,\ \rho(0,.)=\rho_0,\ \rho(1,.)=\rho_1\}. \tag{7}$$

We will detail in the following an algorithm working directly in the set of constraints C. This is crucial for our method since, as it will be proved in the next section, $\mathcal{J}$ has better convexity properties on that set.

2.2. Convexity of the functional in the constraint space. We prove in this section a convexity result for $\mathcal{J}$ on the constraint space, provided some regularity assumptions on the velocity field v. Fortunately, thanks to Hug [29], we have at hand two important regularity results.

Proposition 1 (Hug). *The velocity field $v = \frac{m}{\rho}$ derived from optimal transport belongs to $W^{1,1}(Q)$. Moreover, let $M > 0$ and $\rho_0, \rho_1 \in L^\infty(\Omega)$ of equal mass and such that $0 \leq \rho_0, \rho_1 \leq M$. Then, the time-dependent density ρ, solution of (5-6), is in $L^\infty(Q)$ and $0 \leq \rho \leq M$.*

From now on, we will assume that $\rho_0, \rho_1 \in L^\infty(\Omega)$. The set of constraints can then be restricted to:

$$C_\infty := \{(\rho, m) \in C, \rho \in L^\infty(Q),\ \rho \geq 0 \text{ and } m = \rho v, \quad \text{with } v \in L^1(Q)\},$$

without changing the minimizer.

Proposition 2. *The set C_∞ is non empty and convex.*

Proof. First, C_∞ is non empty. Indeed, because of proposition 1, the solution of the optimal transport problem (ρ, m), for ρ_0, ρ_1 in $L^\infty(\Omega)$ and positive, belongs to C_∞.
We now prove that C_∞ is convex. Let (ρ, m) and (ρ', m') be in C_∞ and $\alpha \in]0, 1[$, then $(\rho_\alpha, m_\alpha) = (\alpha\rho + (1-\alpha)\rho', \alpha m + (1-\alpha)m')$ is in C which is convex and $\rho_\alpha \in L^\infty(Q)$ is positive. Let us now check that $m_\alpha = \rho_\alpha v_\alpha$ with $v_\alpha \in L^1(Q)$.
Because ρ and ρ' are positive, we have, if $\rho_\alpha \neq 0$,

$$\begin{aligned} |v_\alpha| = \left|\frac{m_\alpha}{\rho_\alpha}\right| = \left|\frac{\alpha\ m + (1-\alpha)m'}{\alpha\rho + (1-\alpha)\rho'}\right| &\leq \left|\frac{\alpha m}{\alpha\rho + (1-\alpha)\rho'}\right| + \left|\frac{(1-\alpha)m'}{\alpha\rho + (1-\alpha)\rho'}\right| \\ &\leq \left|\frac{\alpha\ m}{\alpha\rho}\right| + \left|\frac{(1-\alpha)m'}{(1-\alpha)\rho'}\right| = |v| + |v'| \in L^1(Q). \end{aligned}$$

The case $\rho_\alpha = 0$ corresponds to $\rho = \rho' = 0$ which occurs only on a measure zero set since $v = \frac{m}{\rho}, v' = \frac{m'}{\rho'} \in L^1(\Omega)$. Thus v_α is in $L^1(Q)$ and C_∞ is convex. ∎

Using these properties we can derive the following result for the optimal transport problem:

Proposition 3. *The functional $\mathcal{J}$ defined in (5,6) verifies:*

1. *$\mathcal{J}$ is a proper convex lower semicontinuous function on C_∞.*
2. *Let (ρ, m) and (ρ', m') be in C_∞ and $\alpha \in]0, 1[$, such that*

$$\mathcal{J}(\alpha(\rho, m) + (1-\alpha)(\rho', m')) = \alpha\mathcal{J}(\rho, m) + (1-\alpha)\mathcal{J}(\rho', m'), \tag{8}$$

then $\delta\rho = \rho - \rho'$ verifies

$$\begin{cases} \partial_t(\delta\rho) + \nabla_x(w\delta\rho) = 0 \\ \delta\rho|_{\partial Q} = 0, \end{cases} \qquad \text{where} \qquad w = \begin{cases} v \text{ if } \rho > 0 \\ v' \text{ if } \rho' > 0 \\ 0 \text{ otherwise} \end{cases}$$

and $v = v'$ if $\rho\rho' > 0$.

Proof. The first point was proven for example in [4].
To study the convexity of $\mathcal{J}$, let $\alpha \in]0, 1[$, $(\rho, m), (\rho', m') \in \operatorname{dom}\mathcal{J}$ satisfying (8)

$$\int_Q J(\alpha(\rho, m) + (1-\alpha)(\rho', m'))dxdt = \int_Q \left(\alpha J(\rho, m) + (1-\alpha)J(\rho', m')\right) dxdt.$$

Because $m = \rho v$, $m' = \rho' v'$, and $m_\alpha = \rho_\alpha v_\alpha$ with $v, v', v_\alpha \in L^1(Q)$, we have for almost every $(t,x) \in Q$, $J(\rho, m) = \frac{1}{2}\rho v^2 = \frac{|m|^2}{2\rho}$ and so we obtain:

$$\begin{aligned} 0 &= \int_Q \left(\frac{|m_\alpha|^2}{\rho_\alpha} - \alpha \frac{|m|^2}{\rho} - (1-\alpha)\frac{|m'|^2}{\rho'} \right) dxdt \\ &= \int_Q \left(\frac{|m_\alpha|^2 \rho\rho' - \alpha |m|^2 \rho_\alpha \rho' - (1-\alpha)|m'|^2 \rho_\alpha \rho}{\rho_\alpha \rho \rho'} \right) dxdt \\ &= \int_Q \frac{|\alpha\rho v + (1-\alpha)\rho' v'|^2 \rho\rho' - \left(\alpha\rho' |\rho v|^2 + \rho(1-\alpha)|\rho' v'|^2\right)(\alpha\rho + (1-\alpha)\rho')}{(\alpha\rho + (1-\alpha)\rho')\rho\rho'} dxdt \end{aligned}$$

Expanding and collecting terms gives:

$$\int_Q \frac{2\alpha(1-\alpha)\rho^2\rho'^2(v\cdot v') - \alpha(1-\alpha)\rho^2\rho'^2|v|^2 - \alpha(1-\alpha)\rho^2\rho'^2|v'|^2}{(\alpha\rho + (1-\alpha)\rho')\rho\rho'} dxdt = 0,$$

and finally, dividing by $\alpha(1-\alpha) \neq 0$ we obtain:

$$\int_Q \frac{\rho\rho' \left|v - v'\right|^2}{\alpha\rho + (1-\alpha)\rho'} \, dxdt = 0.$$

This leads to

$$\rho\rho' \left|v - v'\right|^2 = 0 \text{ for almost all } t, x \in Q, \tag{9}$$

Thus $v = v'$ if $\rho\rho' > 0$ and we can define

$$w = \begin{cases} v \text{ if } \rho > 0 \\ v' \text{ if } \rho' > 0 \\ 0 \text{ otherwise,} \end{cases}$$

$w \in L^1(Q)$ because $|w|_{L^1} \leq |v|_{L^1} + |v'|_{L^1}$. Moreover, since $(\rho, m), (\rho', m') \in C_\infty$,

$$\begin{aligned} \partial_t\rho + \nabla_x \cdot m = \partial_t\rho + \nabla_x \cdot \rho v = \partial_t\rho + \nabla_x \cdot w\rho \\ \partial_t\rho' + \nabla_x \cdot m' = \partial_t\rho' + \nabla_x \cdot \rho' v' = \partial_t\rho' + \nabla_x \cdot w\rho' \end{aligned} \tag{10}$$

and we obtain the desired result for $\delta\rho = \rho - \rho'$:

$$\partial_t(\delta\rho) + \nabla_x(w\delta\rho) = 0.$$

■

Remark.

- The proposition 3 shows that $\delta\rho = \rho - \rho' \in L^\infty(Q)$ is solution of the continuity equation with homogeneous Dirichlet boundary conditions, associated to the velocity field $w \in L^1(Q)$. As proved in [19, 2], if the velocity field would be in $W^{1,1}(Q)$, there will be a unique solution and thus $\delta\rho = 0$. However for $w \in L^1(Q)$, we are not able to prove the unicity of the solution.
- From another point of view, if we would add the assumption $v \in W^{1,1}(Q)$ in the set C_∞, we could not prove that C_∞ remains convex.

In the following we will only consider positive densities ρ_0, ρ_1 in the space $L^\infty(\Omega)$.

2.3. Reformulation of the problem using the Helmholtz-Hodge decomposition. To work directly in the constraint space C, we use the orthogonal decomposition of $L^2(Q)^{1+n}$, for $n = 1$ or 2, detailed in [23]. Any vector field $v = (\rho, m) \in L^2(Q)^{1+n}$ has the following Helmholtz-Hodge decomposition:

$$(\rho, m) = \nabla \times \phi + \nabla h, \tag{11}$$

where we will denote $\nabla = \nabla_{t,x}$, in the following. Moreover, $h \in H^1(Q)/\mathbb{R}$, and for $n+1 = 2$, $\phi \in H_0^1(Q)$ whereas for $n+1 = 3$, $\phi \in (H_0^1(Q))^3$, and $\nabla \cdot \phi = 0$. Furthermore, if $(\rho, m) \in V(Q)$ (defined in (1)) is divergence-free, the potential h satisfies:

$$\begin{cases} \Delta h = 0 \text{ in } Q, \\ \frac{\partial h}{\partial \nu_Q} = (\rho, m) \cdot \nu_Q \text{ on } \partial Q, \end{cases} \tag{12}$$

where ν_Q is the outward normal of Q and $(\rho, m) \cdot \nu_Q$ represents exactly the boundary conditions in C. In practice, we have first to solve the system (12) to obtain h, which is no more than a Poisson equation with known boundary conditions. Then, knowing h, we have to find the minimum of the new energy

$$E(\phi) = \mathfrak{J}_h(\nabla \times \phi) = \int_Q J\left(\nabla \times \phi(t, x) + \nabla h(t, x)\right)\ dxdt, \tag{13}$$

where J has been defined in (6). Note that, as a composition of an affine operator with a convex function, $\mathfrak{J}_h$ is still convex.

Proposition 4. *In the particular case $n = 1$ (dimension one in space), looking for the minimum in $H_0^1(Q)$ of the energy $E(\phi)$ defined in (13), with the constraint $\rho = \partial_x \phi + \partial_t h > 0$, and h being known, is formally equivalent to solve the equation:*

$$\operatorname{div}_{t,x} \frac{\nabla \phi - \nabla \times h}{|\nabla \phi - \nabla \times h|} = 0. \tag{14}$$

Proof. Since $\mathfrak{J}_h$ is convex, we search for $\phi \in H_0^1(Q)$ cancelling dE:

$$dE(\phi)(\psi) = 0,\ \forall \psi \in H_0^1(Q).$$

$E(\phi) = \int_Q J(\nabla \times \phi + \nabla h) dxdt$, with, by (6), $J(X, Y) = \frac{Y^2}{2X}$ *a.e.*. Then:

$$\partial_X J(X, Y) = -\frac{Y^2}{2X^2} \text{ and } \partial_Y J(X, Y) = \frac{Y}{X}$$

Using $\nabla \times \phi = (\partial_x \phi, -\partial_t \phi)$, the differential of E is given by

$$\begin{aligned} dE(\phi)(\psi) &= \int_Q \left[\partial_X J(\nabla \times \phi + \nabla h)\partial_x \psi - \partial_Y J(\nabla \times \phi + \nabla h)\partial_t \psi\right] dxdt \\ &= \int_Q \left(-\frac{(-\partial_t \phi + \partial_x h)^2}{2(\partial_x \phi + \partial_t h)^2}\partial_x \psi(t, x) - \frac{-\partial_t \phi + \partial_x h}{\partial_x \phi + \partial_t h}\partial_t \psi(t, x)\right) dxdt \\ &= \int_Q \frac{1}{2}\partial_x \left(\frac{(-\partial_t \phi + \partial_x h)^2}{(\partial_x \phi + \partial_t h)^2}\right)\psi + \partial_t \left(\frac{-\partial_t \phi + \partial_x h}{\partial_x \phi + \partial_t h}\right)\psi\ dxdt. \end{aligned}$$

Then $dE(\phi)(\psi) = 0$ for all $\psi \in H_0^1(Q)$ if and only if

$$\partial_t \left(\frac{-\partial_t \phi + \partial_x h}{\partial_x \phi + \partial_t h} \right) + \frac{1}{2} \partial_x \left(\frac{(-\partial_t \phi + \partial_x h)^2}{(\partial_x \phi + \partial_t h)^2} \right) = 0. \tag{15}$$

If we now denote

$$\begin{pmatrix} v \\ u \end{pmatrix} = \nabla \times \phi + \nabla h = \begin{pmatrix} \partial_x \phi + \partial_t h \\ -\partial_t \phi + \partial_x h \end{pmatrix},$$

equation (15) rewrites:

$$\begin{aligned} 0 = \partial_t \left(\frac{u}{v} \right) + \frac{1}{2} \partial_x \left(\frac{u}{v} \right)^2 &= \frac{(\partial_t u) v - u \partial_t v}{v^2} + \frac{u}{v} \frac{(\partial_x u) v - u \partial_x v}{v^2} \\ &= \frac{(\partial_t u) v^2 - uv \partial_t v + uv \partial_x u - u^2 \partial_x v}{v^3}. \end{aligned}$$

Remarking that:

$$\nabla \phi - \nabla \times h = \begin{pmatrix} -u \\ v \end{pmatrix} \quad \text{and} \quad |\nabla \phi - \nabla \times h|^2 = u^2 + v^2,$$

we compute:

$$\begin{aligned} \operatorname{div}_{t,x} \left(\frac{(-u, v)}{(u^2 + v^2)^{1/2}} \right) &= \partial_t \left(\frac{-u}{(u^2 + v^2)^{1/2}} \right) + \partial_x \left(\frac{v}{(u^2 + v^2)^{1/2}} \right) \\ &= -\frac{\partial_t u (u^2 + v^2) - u^2 \partial_t u - uv \partial_t v}{(u^2 + v^2)^{3/2}} + \frac{\partial_x v (u^2 + v^2) - vu \partial_x u - v^2 \partial_x v}{(u^2 + v^2)^{3/2}} \\ &= \frac{-(\partial_t u) v^2 + uv \partial_t v + (\partial_x v) u^2 - vu \partial_x u}{(u^2 + v^2)^{3/2}} \\ &= \left(\partial_t \left(\frac{u}{v} \right) + \frac{1}{2} \partial_x \left(\frac{u}{v} \right)^2 \right) \left(\frac{-v^3}{(u^2 + v^2)^{3/2}} \right). \end{aligned} \tag{16}$$

Since we assumed $v = \partial_x \phi + \partial_t h > 0$, (15) is equivalent to

$$\operatorname{div}_{t,x} \left(\frac{(-(-\partial_t \phi + \partial_x h), \partial_x \phi + \partial_t h)}{((-\partial_t \phi + \partial_x h)^2 + (\partial_x \phi + \partial_t h)^2)^{1/2}} \right) = 0.$$

which is the expected equation (14). ∎

Remarks.

1. The result of proposition 4 stems from different facts. First, let us observe that in dimension two the Hessians H_J and $H_{\overline{J}}$ of respectively the functionals $J(X, Y) = \frac{Y^2}{2X}$ and $\overline{J}(X, Y) = \sqrt{X^2 + Y^2}$ are proportional, which is no more true in higher dimensions. Indeed,

$$H_J = \begin{pmatrix} \partial_{XX} J & \partial_{XY} J \\ \partial_{XY} J & \partial_{YY} J \end{pmatrix} = \frac{1}{X^3} \begin{pmatrix} Y^2 & -XY \\ -XY & X^2 \end{pmatrix}$$

and

$$H_{\overline{J}} = \frac{1}{(X^2+Y^2)^{3/2}} \begin{pmatrix} Y^2 & -XY \\ -XY & X^2 \end{pmatrix}$$

Now, define

$$\overline{E}(\phi) = \int_Q \overline{J}\left(\nabla \times \phi(t,x) + \nabla h(t,x)\right) dxdt = \int_Q \left|\nabla \times \phi(t,x) + \nabla h(t,x)\right| \, dxdt.$$

Then dE can be rewritten in terms of $\overline{E}$. For all $\psi \in H_0^1(Q)$, using anew the notation $(v,u) = \nabla \times \phi + \nabla h$,

$$\begin{aligned}
dE(\phi)(\psi) &= \int_Q \left[\partial_X J(v(t,x),u(t,x))\partial_x\psi - \partial_Y J(v(t,x),u(t,x))\partial_t\psi\right] \, dxdt \\
&= -\int_Q \left[\partial_x\partial_X J(v,u) - \partial_t\partial_Y J(v,u)\right]\psi \, dxdt \\
&= -\int_Q [\partial_x v \partial_{XX} J(v,u) + \partial_x u \partial_{XY} J(v,u) \\
&\quad -\partial_t v \partial_{XY} J(v,u) - \partial_t u \partial_{YY} J(v,u)]\psi \, dxdt \\
&= -\int_Q H_J(v,u) : \begin{pmatrix} \partial_x v & \partial_x u \\ -\partial_t v & -\partial_t u \end{pmatrix} \psi \, dxdt \\
&= -\int_Q \frac{(v^2+u^2)^{3/2}}{v^3} H_{\overline{J}}(v,u) : \begin{pmatrix} \partial_x v & \partial_x u \\ -\partial_t v & -\partial_t u \end{pmatrix} \psi \, dxdt.
\end{aligned}$$

where $'' :''$ denotes the euclidean dot product between two matrices.
Because $v^2+u^2 = |\nabla \times \phi + \nabla h|^2$ is non zero, we obtain:

$$dE(\phi)(\psi) = 0 \iff d\overline{E}(\phi)(\psi) = 0$$

Moreover, the norms $|\nabla\phi - \nabla \times h|$ and $|\nabla \times \phi + \nabla h|$ are equal, so minimizing

$$\int_Q |\nabla \times \phi(t,x) + \nabla h(t,x)| dxdt$$

is equivalent to minimizing

$$\int_Q |\nabla \phi(t,x) - \nabla \times h(t,x)| dxdt$$

which Euler-Lagrange equation is (14).

2. The minimisation of this new functional is simpler than the minimization of $\mathcal{J}$ as we will see in section 3.
3. The L^2 orthogonality of $\nabla\phi$ and $\nabla \times h$ provided by the decomposition (11) and the orthogonality of $\nabla\phi - \nabla \times h$ and $\nabla \times \phi + \nabla h$ allowed to simplify the calculation (16).

4. This proposition does not hold in the 2D+t case because -in particular- H_J and $H_{\bar{J}}$ are no more proportional.

In conclusion, in 1D, searching for the minimum of E in (13) amounts to the resolution of a minimal surface equation on each level set of the potential ϕ equipped with appropriate Dirichlet boundary conditions. Finally we obtain the following systems:

$$\left\{ \begin{array}{l} \Delta h = 0 \text{ in } Q, \\ \frac{\partial h}{\partial \nu_Q} = (\rho, m) \cdot \nu_Q \text{ on } \partial Q \end{array} \right. \tag{17}$$

and

$$\left\{ \begin{array}{l} \operatorname{div}_{t,x} \frac{\nabla\phi - \nabla\times h}{|\nabla\phi - \nabla\times h|} = 0, \\ \phi = 0 \text{ on } \partial Q. \end{array} \right. \tag{18}$$

The equation (18) is the Euler-Lagrange equation of the functional (convex as the composition of a linear operator with a convex function)

$$\mathcal{J}_{ms}(\nabla\phi) := \int_Q |\nabla\phi - \nabla \times h| \; dxdt = ||\nabla\phi - \nabla \times h||_1. \tag{19}$$

Section 3.1 will detail the minimization of this new functional $\mathcal{J}_{ms}$, instead of directly solving the equation (18).

Remark. Optimal transport between two densities applies under a condition of iso-mass of those densities. However in applications one might want to interpolate between to images which are not of equal mass. This case referred as to unbalanced optimal transport has been soon addressed [5, 39, 34, 14, 15]. Our present formulation could be generalized to that context by incorporating a source term in the Laplace equation in h.

3. Two algorithms using a first order primal-dual formulation. Let now denote $\mathcal{X}$ and $\mathcal{Y}$ two real vector spaces equipped with the same inner product $\langle ., . \rangle$ and norm $|.| = \langle ., . \rangle^{1/2}$. Our formulation of the optimal problem (13) can be viewed as searching the minimum of a functional of the form:

$$\min_{\phi} \; F(K(\phi)) + \iota_{C_0}(\phi), \tag{20}$$

where $F \in \Gamma_0(\mathcal{Y})$, the set of proper, convex, lower semi-continuous (l.s.c.) applications $\mathcal{Y} \to \mathbb{R}_+$, $K : \mathcal{X} \to \mathcal{Y}$ is a linear continuous operator and ι_{C_0} is the indicator function of the set $C_0 := \{\phi = 0 \text{ on } \partial Q\}$ which is in $\Gamma_0(\mathcal{Y})$. Such minimization problem falls into the framework of Chambolle and Pock [12], who solve (20) using its primal dual formulation (see [41]):

$$\min_{\phi} \max_{z} \langle K\phi, z \rangle + \iota_{C_0}(\phi) - F^*(z). \tag{21}$$

F^* is the Legendre transform of F (see [4]), and is defined by

$$F^* : \left\{ \begin{array}{ll} \mathcal{Y} & \to [0, +\infty) \\ y & \mapsto \max_{x} \langle x, y \rangle - F(y). \end{array} \right.$$

The proximal operator of the function F^* defined, for $\gamma > 0$, by

$$\mathrm{prox}_{\gamma F^*} : \left\{ \begin{array}{lll} \mathcal{Y} & \to & \mathcal{Y} \\ x & \mapsto & \underset{y}{\mathrm{argmin}} \left(\frac{1}{2}|x-y|^2 + \gamma F^*(y) \right), \end{array} \right.$$

and the norm of the operator K, $||K|| = \sup\{|Kx| : x \in \mathcal{X}, |x| \le 1\}$, are used in the primal-dual algorithm, summarized as follows.

Algorithm 1 (General Chambolle-Pock).
Initialization: $\tau, \sigma > 0, \theta \in [0,1]$, $(\phi^0, z^0 = K\phi^0, \tilde{\phi}^0 = \phi^0)$.
Iterations:

$$\begin{aligned} z^{i+1} &= \mathrm{prox}_{F^*}(z^i + \sigma(K\tilde{\phi}^i)) \\ \phi^{i+1} &= \mathrm{prox}_{\iota_{C_0}}(\phi^i - \tau K^* z^{i+1}) \\ \tilde{\phi}^{i+1} &= \phi^{i+1} + \theta(\phi^{i+1} - \phi^i). \end{aligned}$$

Because the 1D case leads to a minimal surface formulation (see section 2.3), we will present two different algorithms. The first one solves equation (14) minimizing the functional $\mathcal{J}_{ms}$ defined in (19), whereas the second one minimizes the functional $\mathcal{J}_h$ defined in (13).

Convergence of the algorithm: Assuming that $\mathcal{X}$ and $\mathcal{Y}$ have finite dimension, and that the problem (21) has a solution $(\hat{\phi}, \hat{z})$, it has been proved in [11, 12, 25] that for $\theta = 1$ and $\sigma\tau||K||^2 < 1$, the sequence (ϕ^i, z^i) computed with Algorithm 1, converges to the exact solution $(\hat{\phi}, \hat{z})$, for any initial condition ϕ^0.
We thus have the convergence of the algorithm for the discretized problem which will be now detailed in both cases.

3.1. Primal dual algorithm for the minimal surface equation.

We first detail the algorithm when considering the minimal surface equation (14), which consists to minimize $\mathcal{J}_{ms}$ defined in (19), with suitable boundary conditions:

$$\min ||\nabla\phi - \nabla \times h||_1 + \iota_{C_0}(\phi) = \min\ \mathcal{J}_{ms}(\nabla\phi) + \iota_{C_0}(\phi) \tag{22}$$

Here we took $K = \nabla$, the gradient operator (linear continuous operator from $\mathcal{X} = H^1(Q)$ to $\mathcal{Y} = (L^2(Q))^2$) and $F = \mathcal{J}_{ms}(y) = ||y - \nabla \times h||_1$ which is proper, convex and continuous. The primal-dual formulation of this primal problem is

$$\min_{\phi} \max_{z} \langle z.\nabla\phi\rangle + \iota_{C_0}(\phi) - \mathcal{J}^*_{ms}(z).$$

Stating the problem in a discrete setting, $||.||_1$ writes, on a discrete centered bidimensional grid G^c which will be defined in 4.1: for $x = (x_k)_{k\in G^c}$, $x_k \in \mathbb{R}^2$,

$$||x||_1 = \sum_{k\in G^c} |x_k| \text{ where } x_k = (x^1, x^2), \quad |x_k| = \sqrt{(x^1_k)^2 + (x^2_k)^2},$$

and we have,

$$\mathrm{prox}_{\gamma||.||_1}(x) = (\mathrm{prox}_{\gamma|.|}\, x_k)_{k\in G^c}.$$

For $(a,b)=((a_k,b_k))_{k\in G^c}$ and $(\nabla\times h_k)_{k\in G^c}$ the values of respectively $\nabla\phi$ and $\nabla\times h$ on the grid G^c,

$$\mathcal{J}_{ms}(a,b)=||(a-\partial_x h, b+\partial_t h)||_1=\sum_{k\in G^c} J_{ms}(a_k,b_k)=\sum_{k\in G^c}|(a_k-\partial_x h_k, b_k+\partial_t h_k)|$$

where $J_{ms}(a_k,b_k)=|(a_k-\partial_x h_k, b_k+\partial_t h_k)|$, $\nabla\times h_k$ being known. The following proposition holds.

Proposition 5. *For all c, vector valued on the grid G^c, one has:*

$$\text{prox}_{\gamma\mathcal{J}^*_{ms}}(c)=\min(c-\gamma\nabla\times h, \frac{(c-\gamma\nabla\times h)}{|c-\gamma\nabla\times h|}),$$

Then we can observe that $\text{prox}_{\gamma\mathcal{J}^*_{ms}}$ is a pointwise Euclidean projection onto a unit ball in $\mathbb{R}^{2p}$ where p is the size of the grid.

Proof. See appendix A ■

Moreover, since ι_{C_0} is the indicator function on a closed, non empty convex set, its proximal operator is the projection onto the set C_0 (see [17]), which we will denote by $\mathcal{P}_{C_0}$.
Its computation merely corresponds to set the boundary values of ϕ to zero.
Finally, in this case we obtain the following algorithm:

Algorithm 2 (PDHHMS).
Initialization: $\tau,\sigma>0,\theta\in[0,1]$, $(\phi^0, z^0=\nabla\phi^0, \bar{\phi}^0=\phi^0)$.
Iterations:

$$\begin{aligned}
z^{i+1}&=\text{prox}_{\mathcal{J}^*_{ms}}(z^i+\sigma(\nabla\bar{\phi}^i-\nabla\times h))\\
\phi^{i+1}&=\mathcal{P}_{C_0}(\phi^i-\tau\nabla.z^{i+1})\\
\bar{\phi}^{i+1}&=\phi^{i+1}+\theta(\phi^{i+1}-\phi^i).
\end{aligned}$$

It has been shown in [11] that for $\theta=1$ and $\sigma\tau||\nabla||^2<1$, ϕ^i computed with the above algorithm, converges to the solution of the discrete version of problem (22):

$$\min_{\phi}\sum_{k\in G^c} J_{ms}(\nabla\phi_k)+\iota_{C_0}(\phi_k) \tag{23}$$

for any choice of ϕ^0.

3.2. Primal dual algorithm for the functional $\mathcal{J}_h$. We now want to minimize (13) whose primal-dual formulation is:

$$\min_{\phi}\max_{z}\langle K\phi, z\rangle+\iota_{C_0}(\phi)-\mathcal{J}^*_h(z), \tag{24}$$

with $K=\nabla\times$, the curl operator (linear continuous operator from $(H^1(Q))^3$ to $(L^2(Q))^3$) and $\mathcal{J}^*_h:(L^2(Q))^3\to[0,+\infty)$ is a proper, convex, lower semicontinuous function. The discrete objective functional $\mathcal{J}$ reads for (ρ, m) defined on the centered tridimensional grid G^c (defined in section 4.1):

$$\mathcal{J}(\rho, m) = \sum_{k \in G^c} J(\rho_k, m_k), \tag{25}$$

where the functional J is defined in (6), and then,

$$\text{prox}_{\gamma \mathcal{J}}(x) = (\text{prox}_{\gamma J}(x_k))_{k \in G^c}.$$

As proved in [6], the Legendre transform of $\mathcal{J}$ is the indicator function of a convex set, $\mathcal{J}^* = i_{P_\mathcal{J}}$ where

$$\begin{cases} P_\mathcal{J} = \{(z_1, z_2);\ \forall k \in G^c,\ (z_1, z_2)_k \in P_J^n\} \\ P_J^n = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n,\ t + \frac{|x|^2}{2} \leq 0\}. \end{cases}$$

This implies that $\text{prox}_{\gamma J^*}$ is the projection onto the paraboloid P_J^n, which we will denote by $\mathcal{P}_{P_J^n}$. This is a consequence of the fact that $\mathcal{J}^*$ is proper, convex, lower semi-continuous and 1-homogeneous.

Proposition 6. *The projection onto the paraboloid $P_J^2 = \{(a, b) \in \mathbb{R} \times \mathbb{R}^2,\ a + \frac{|b|^2}{2} = 0\}$ of a point $(z_1, z_2) \in \mathbb{R} \times \mathbb{R}^2$ outside of P_J^2 is*

$$\mathcal{P}_{P_J^2}(z_1, z_2) = \left(\frac{-\beta^2}{2}, \frac{z_2}{1 + z_1 + \beta^2/2} \right)$$

where β is the real positive solution of the cubic equation:

$$\frac{-X^3}{2} - (1 + z_1)X + |z_2| = 0\ .$$

In order to prove this proposition we begin by studying the one dimensional case in space.

Proposition 7. *The projection onto the paraboloid $P_J^1 = \{(a, b) \in \mathbb{R} \times \mathbb{R},\ a + \frac{b^2}{2} = 0\}$ is*

$$\mathcal{P}_{P_J^1}(z_1, z_2) = \left(\frac{-\beta^2}{2}, \beta \right)$$

where β is the real solution of the cubic equation, whose sign is the same as the sign of z_2

$$\frac{-X^3}{2} - (1 + z_1)X + z_2 = 0\ .$$

Proof. To explicit the projection onto the paraboloid P_J^1, we introduce:

$$(\alpha, \beta) = \mathcal{P}_{P_J^1}(z_1, z_2)$$

Remarking that the normal to the paraboloid at the point (α, β) is the vector $(1, \beta)$, we obtain the following equation:

$$(\alpha, \beta) - (z_1, z_2) \ // \ (1, \beta) \Leftrightarrow (\alpha - z_1)\beta - (\beta - z_2) = 0.$$

Then we have the following system

$$\begin{cases} (\alpha - z_1)\beta - (\beta - z_2) & = 0 \\ \alpha + \frac{\beta^2}{2} & = 0 \end{cases} \Leftrightarrow \begin{cases} \frac{-\beta^3}{2} - (1 + z_1)\beta + z_2 & = 0 \\ \alpha & = -\frac{\beta^2}{2} \end{cases}$$

To obtain the projection we have to solve the cubic equation in (5) which is done by using Cardano's method. The solution of the equation we are interested in is the real one whose sign is the same as the sign of z_2. ■

We are then able to prove proposition 6.

Proof. Let $(\tilde{a}, \tilde{b})$ be the projection of (z_1, z_2) onto the paraboloid P_J^2. Since (z_1, z_2) is on the normal to the paraboloid at $(\tilde{a}, \tilde{b})$, of orientation vector $(1, \tilde{b})$ we have:

$$(z_1 - \tilde{a}, z_2 - \tilde{b}) = \lambda(1, \tilde{b}). \tag{26}$$

so $\lambda = z_1 - \tilde{a}$. To find the projection in 3D we just have to project in 2D $(z_1, |z_2|)$ onto the paraboloid P_J^1, which corresponds to the intersection between the paraboloid P_J^2 in 3D and the plan which contains the time (z_1)-axis and (z_1, z_2). Thus we obtain

$$(\tilde{a}, \tilde{b}) = \mathcal{P}_{P_J^2}(z_1, z_2) = \mathcal{P}_{P_J^1}(z_1, |z_2|) = \left(\frac{-\beta^2}{2}, \beta \right)$$

where β is the real positive solution of the cubic equation : $\frac{-X^3}{2} - (1 + z_1)X + |z_2| = 0$. Then $\tilde{a} = -\beta^2/2$ and equation (26) leads to:

$$\lambda = z_1 - \tilde{a} = z_1 + \beta^2/2 \text{ and } \tilde{b} = \frac{z_2}{1 + \lambda} = \frac{z_2}{1 + z_1 + \beta^2/2}.$$

Let now check that $1 + \lambda \neq 0$. If $1 + \lambda = 0$ then, from (26): $z_2 - \tilde{b} = -\tilde{b}$ and so $z_2 = 0$. The pair (z_1, z_2) is outside of P_J^2 so $z_1 + |z_2|^2/2 > 0$ and in this case $z_1 > 0$. On the other hand, the pair $(\tilde{a}, \tilde{b})$ is in P_J^2 so $\tilde{a} \leq 0$. But from (26) we have

$$z_1 = \tilde{a} + \lambda = \tilde{a} - 1 < 0$$

which is not possible. Thus $1 + \lambda \neq 0$. ■

Remark. Proposition 6 provides an exact and straightforward formula, usefull for the computation of the proximal operator.

Now, looking at the functional $\mathcal{J}_h$, we have for $(a, b) = \nabla \times \phi$:

$$\mathcal{J}_h(a, b) = \mathcal{J}(a + \partial_t h, b + \nabla_x h) = \sum_{k \in G^c} J_h(a_k, b_k) = \sum_{k \in G^c} J(a_k + \partial_t h_k, b_k + \nabla_x h_k).$$

with $J_h(a_k, b_k) = J(a_k + \partial_t h_k, b_k + \nabla_x h_k)$. This enables us to deduce from J^* the form of J_h^* and the form of $\text{prox}_{\gamma J_h^*}$ from the one of $\text{prox}_{\gamma J^*}$. If we denote $c = (a, b)$ we have the following proposition:

Proposition 8. *One has for all $c \in \mathbb{R}^{1+n}$*

$$J_h^*(c) = J^*(c) - \langle \nabla h, c \rangle, \quad \textit{and} \quad \text{prox}_{\gamma J_h^*}(c) = \text{prox}_{\gamma J^*}(c - \gamma \nabla h).$$

Proof. By definition of the Legendre transform:

$$\begin{aligned}
J_h^*(c) &= \max_x \langle x, c \rangle - J_h(x) \\
&= \max_x \langle x, c \rangle - J(x + \nabla h) \\
&= \max_x \langle x - \nabla h, c \rangle - J(x) \\
&= J^*(c) - \langle \nabla h, c \rangle.
\end{aligned}$$

The proximal operator is given by:

$$\begin{aligned}
\text{prox}_{\gamma J_h^*}(c) &= \underset{x}{\text{argmin}} \frac{1}{2}|x - c|^2 + \gamma J_h^*(x) \\
&= \underset{x}{\text{argmin}} \frac{1}{2}|x - c|^2 + \gamma(J^*(x) - \langle \nabla h, x \rangle) \\
&= \underset{x}{\text{argmin}} \frac{1}{2}|x - \gamma \nabla h - c|^2 + \gamma J^*(x) \\
&= \text{prox}_{\gamma J^*}(c + \gamma \nabla h).
\end{aligned}$$

■

Finally, the primal-dual algorithm in our case leads to the PDHH-algorithm.

Algorithm 3 (PDHH).
Initialization: $\tau, \sigma > 0, \theta \in [0, 1]$, $(\phi^0, z^0 = \nabla \times \phi^0, \tilde{\phi}^0 = \phi^0)$.
Iterations:

$$\begin{aligned}
z^{i+1} &= \mathcal{P}_{P_J^2}(z^i + \sigma(\nabla \times \tilde{\phi}^i + \nabla h)) \\
\phi^{i+1} &= \mathcal{P}_{C_0}(\phi^i - \tau \nabla^* \times z^{i+1}) \\
\tilde{\phi}^{i+1} &= \phi^{i+1} + \theta(\phi^{i+1} - \phi^i).
\end{aligned}$$

As before, for $\theta = 1$ and $\sigma\tau||K||^2 < 1$, ϕ^i computed with the above algorithm converges to the solution of the discrete version of problem (21). The computation of $\mathcal{P}_{P_J^2}$ amounts to solving a cubic equation by grid point, while $\mathcal{P}_{C_0}$ merely corresponds to setting the boundary condition to zero.

Remark. We can observe that analytically the curl operator is self-adjoint but since the discrete curl operator depends on the discrete derivative, it might be not self-adjoint. That is why we will keep both notations ∇ and ∇^*.

Nonlinear case. In [44], Valkonen proposes to extend the primal-dual algorithm to the case where $K \in \mathcal{C}^2$ is allowed to be nonlinear. Introducing this idea in the above algorithm PDHH leads to: instead of minimizing $\mathcal{J}_h(K(\phi))$ with $K(\phi) = \nabla \times \phi$, we minimize $\mathcal{J}(K(\phi))$ where $K(\phi) = \nabla \times \phi + \nabla h$. The extended algorithm reads:

Algorithm 4.
Initialization: $\tau, \sigma, \theta,\ (\phi^0, z^0 = K\phi^0)$.
Iterations:

$$\begin{aligned}
\phi^{i+1} &:= \mathrm{prox}_{\tau \iota_{C_0}}(\phi^i - \tau[\nabla K(\phi^i)]^* z^i),\\
\tilde{\phi}^{i+1} &:= \phi^{i+1} + \theta(\phi^{i+1} - \phi^i),\\
z^{i+1} &:= \mathrm{prox}_{\sigma \mathcal{J}^*}(z^i + \sigma K(\tilde{\phi}^{i+1})),
\end{aligned}$$

where $[\nabla K]$ is the linearization of the operator K. Our operator being affine, its linearization is just $[\nabla K] = \nabla\times$. Thus this algorithm for nonlinear operator K is in our case equivalent to Algorithm 3 (PDHH).

4. Numerical applications and comparisons.

In this section we first present the discrete setting and then compare our algorithm to state-of-art methods.

4.1. Discrete setting.

The discretization on uniform grids follows the discretization method introduced in [36] and uses uniform staggered grids as in fluid dynamics.

4.1.1. Discrete settings for 1D+t images.

We first describe the discrete grids used in the computations for 1D+t images, which are the same when minimizing the functionals (13) and (22).

Centered grid. The evaluation of the dual variable $z(t,x)$ is done on a regular grid G^{c1} of size $M \times N$, whereas the one of the primal variable $\phi(t,x)$ is done on a regular grid G^{c2} of size $(M+1)\times(N+1)$. These regular grids are defined by

$$\begin{aligned}
G^{c1} &= \{t_i, x_j\}_{1\le i\le M,\ 1\le j\le N},\\
G^{c2} &= \{t_{i-1/2}, x_{j-1/2}\}_{1\le i\le M+1,\ 1\le j\le N+1},
\end{aligned}$$

with $t_i = \frac{i}{M+1},\ x_j = \frac{j}{N+1}$ the discrete locations in the domain $Q = (0,1)\times(0,1)$.

Staggered grid. We now introduce the grid G^{s1}, which provides a discretization coherent with the divergence of (ρ, m) and which is defined by:

$$\begin{aligned}
G^{s1}_t &= \{t_{i-1/2},\ x_j\}_{1\le i\le M+1,\ 1\le j\le N},\\
G^{s1}_x &= \{t_i,\ x_{j-1/2}\}_{1\le i\le M,\ 1\le j\le N+1}.
\end{aligned}$$

Interpolation operator. To evaluate the values of a vector $u = (u^1, u^2)$ on the centered grid G^{c1}, from the knowledge of its values $\bar{u}$ on the staggered grid G^{s1}, we need an interpolation operator, whose first component is taken equal to:

$$\forall\ 1\le i\le M,\ \forall\ 1\le j\le N,\quad u^1_{i,j} = \frac{1}{2}(\bar{u}^1_{i+1/2,j} + \bar{u}^1_{i-1/2,j}),$$

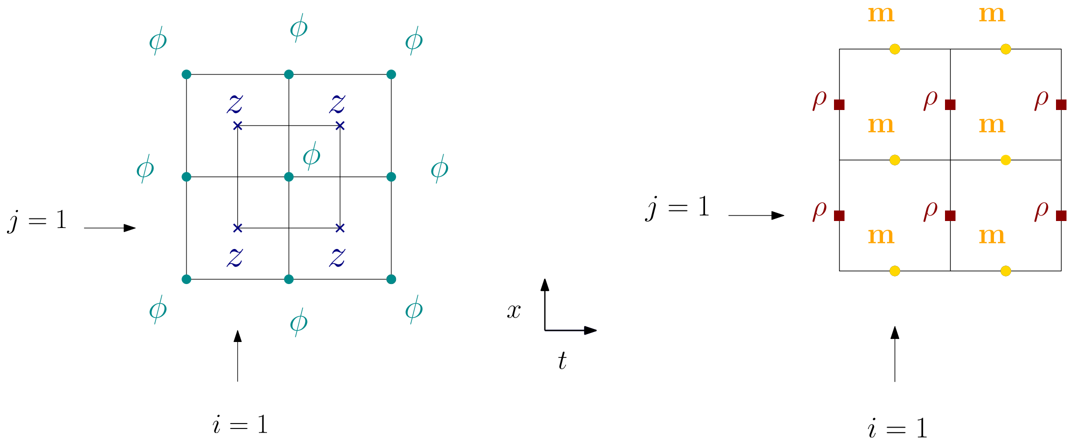


Figure 1: Grids for 1D+t images.

and its adjoint operator to go from $v \in G^{c1}$ to $\bar{v} \in G^{s1}$:

$$\bar{v}^1_{i-1/2,j} = \begin{cases} v_{1,j} & \text{if } i = 1 \\ v_{i,j} + v_{i-1,j} & \text{if } 2 \leq i \leq M \\ v_{M,j} & \text{if } i = M+1. \end{cases}$$

Curl, gradient and divergence operators. The curl, gradient and divergence operators are discretized by using centered finite differences. The **discrete gradient**, which is a vector of matrices, and the divergence operator, which is its adjoint, are

$$\nabla = \begin{pmatrix} \partial_t \\ \partial_x \end{pmatrix} \text{ and } \nabla^*. = \begin{pmatrix} \partial_t^* \\ \partial_x^* \end{pmatrix}. = \partial_t^* + \partial_x^*.$$

The discrete partial derivative operator with respect to the first component reads: $\partial_t : G^{c1} \to G^{s1}_t$, and for $v \in G^{c1}$:

$$(\partial_t v)_{i+1/2,j} = v_{i+1,j} - v_{i,j}, \quad \forall 1 \leq i \leq M-1, \forall 1 \leq j \leq N.$$

The adjoint partial derivative operator for $\bar{u} = (\bar{u}^1, \bar{u}^2) \in G^{s1}$ is defined by

$$(\partial_t^* \bar{v})_{i,j} = \begin{cases} -\bar{v}^1_{1+1/2,j} & \text{if } i = 1 \\ \bar{v}^1_{i-1/2,j} - \bar{v}^1_{i+1/2,j} & \text{if } 2 \leq i \leq M-1 \\ \bar{v}^1_{M-1/2,j} & \text{if } i = M. \end{cases}$$

Regarding the **curl** operator and its adjoint, we consider:

$$\nabla\times = \begin{pmatrix} \partial_x \\ -\partial_t \end{pmatrix} \text{ and } \nabla^*\times = \begin{pmatrix} \partial_x^* \\ -\partial_t^* \end{pmatrix} = \partial_x^* - \partial_t^*.$$

On the second centered grid G^{c2}, the discrete partial derivative operator with respect to the first component reads: $\partial_t : G^{c2} \to G_x^{s1}$, and for $\tilde{v} \in G^{c2}$:

$$(\partial_t \tilde{v})_{i,j-1/2} = \tilde{v}_{i+1/2,j-1/2} - \tilde{v}_{i-1/2,j-1/2}, \quad 1 \leq i \leq M,\ 1 \leq j \leq N+1,$$

and the adjoint partial derivative operator for $\bar{u} = (\bar{u}^1, \bar{u}^2) \in G^{s1}$ is defined by

$$(\partial_t^* \bar{u}^2)_{i-1/2,j-1/2} = \begin{cases} -\bar{u}^2_{1,j-1/2} & \text{if } i = 1 \\ \bar{u}^2_{i-1,j-1/2} - \bar{u}^2_{i,j-1/2} & \text{if } 2 \leq i \leq M \\ \bar{u}^2_{M,j-1/2} & \text{if } i = M+1. \end{cases}$$

4.1.2. Discrete settings for 2D+t images. We now describe the discrete grids used in the computations for 2D+t images.

Centered grid. The regular grid

$$G^c = \{t_i, x_j, y_k\}_{1 \leq i \leq M,\ 1 \leq j \leq N,\ 1 \leq k \leq P},$$

with $t_i = \frac{i}{M}$, $x_j = \frac{j}{N}$, $y_k = \frac{k}{P}$ the discrete locations in the domain Q, is used to evaluate $\rho(t,x,y)$ and $m(t,x,y)$, to calculate the functional, and to evaluate the dual variable z.

Staggered grid. We introduce two staggered grids to evaluate the divergence and the curl operators. The grid G^{s1} provides a discretization coherent with the divergence of (ρ, m) and is defined by:

$$\begin{aligned} G_t^{s1} &= \{t_{i-1/2},\ x_j,\ y_k\}_{1 \leq i \leq M+1,\ 1 \leq j \leq N,\ 1 \leq k \leq P}, \\ G_x^{s1} &= \{t_i,\ x_{j-1/2},\ y_k\}_{1 \leq i \leq M,\ 1 \leq j \leq N+1,\ 1 \leq k \leq P}, \\ G_y^{s1} &= \{t_i,\ x_j,\ y_{k-1/2}\}_{1 \leq i \leq M,\ 1 \leq j \leq N,\ 1 \leq k \leq P+1}. \end{aligned}$$

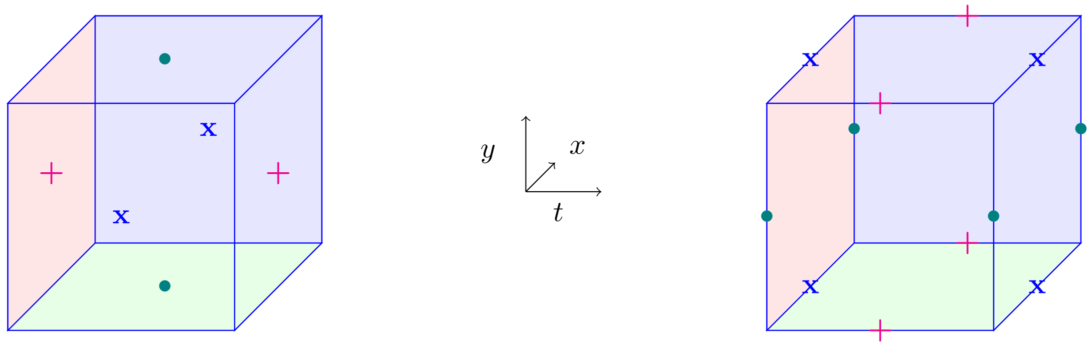
Our staggered grid G^{s_2} is used to evaluate ϕ such that $\nabla \times \phi$, because $(\rho, m) = \nabla \times \phi + \nabla h$, leaves on the staggered grid G^{s_1}:

$$\begin{aligned} G_t^{s2} &= \{t_i,\ x_{j-1/2},\ y_{k-1/2}\}_{1 \leq i \leq M,\ 1 \leq j \leq N+1,\ 1 \leq k \leq P+1}, \\ G_x^{s2} &= \{t_{i-1/2},\ x_j,\ y_{k-1/2}\}_{1 \leq i \leq M+1,\ 1 \leq j \leq N,\ 1 \leq k \leq P+1}, \\ G_y^{s2} &= \{t_{i-1/2},\ x_{j-1/2},\ y_k\}_{1 \leq i \leq M+1,\ 1 \leq j \leq N+1,\ 1 \leq k \leq P}. \end{aligned}$$

Interpolation, gradient and divergence operators. These operators are the same as those described for 1D+t images.

Curl operators. Regarding the curl operator, in order to use the primal dual algorithm, we need to define the discrete adjoint operator of the curl. Because the curl operator is given by the following matrix

$$\nabla \times = \begin{pmatrix} 0 & -\partial_y & \partial_x \\ \partial_y & 0 & -\partial_t \\ -\partial_x & \partial_t & 0 \end{pmatrix}$$



A grid G^{s_1} to evaluate the continuity equation for $(\rho, m_1, m_2) = \nabla \times \phi + \nabla h$.

A grid G^{s_2} to define (ϕ_1, ϕ_2, ϕ_3) whose curl lives on the staggered grid G^{s_1}.

Figure 2: Staggered grids for 2D+t images

the appropriate adjoint curl operator has to be the opposite of the curl derived from the adjoint partial derivative operators:

$$\nabla^* \times = \begin{pmatrix} 0 & \partial_y^* & -\partial_x^* \\ -\partial_y^* & 0 & \partial_t^* \\ \partial_x^* & -\partial_t^* & 0 \end{pmatrix}.$$

4.2. Numerical tests and comparisons. In order to evaluate the performances of our methods, we need to define the choice of the parameter σ used in the first step of the primal-dual algorithm. To that end, we first have to compute a reference solution (ρ_s, m_s) of the discrete problem by computing 10^6 iterations. We then choose σ such that the errors on m and ρ are minimal after a given number of iterations ($<< 10^6$) for our algorithm (PDHH, alg. 3). We also take the same σ for the minimal surface formulation (PDHHMS, alg. 2) and for the algorithm developed in [36] that we will denote PDPOP in the following: more precisely, PDPOP^{gh} will denote the code available on github[1], whereas PDPOP will denote the same method where we modified the computation of the proximal operator of $\mathcal{J}^*$, to be more efficient, as seen later.

All our code is available for download on GitHub[2]. The σ we obtain depends on the initialization, the densities ρ_0 and ρ_1, and the chosen number of iterations. The algorithm (modified PDPOP of [36]) to which we compare is the resolution of the dynamic optimal transport using the primal-dual algorithm. The only difference between this algorithm and the PDHH algorithm is the decomposition $(\rho, m) = \nabla \times \phi + \nabla h$, which allows to remain in the divergence-free space, and therefore the second step of the algorithm which is the projection onto the divergence-free constraint and the boundary conditions. In PDHH, we only need the projection onto the boundary conditions.

All our computations have been performed on Intel Core i7 (Dual core, 2.8GHz).

[1] https://github.com/gpeyre/2013-SIIMS-ot-splitting
[2] https://github.com/MorganeMartinHenry/Primal-dual-formulation-optimal-transport-Helmholtz-Hodge-decomposition

4.2.1. Algorithms for 1D+t images. In the following we will use the parameters $\sigma = 1$, $\tau = 0.99/L\sigma$ and $\theta = 1$. We chose σ such that the errors on m and ρ are minimal after 50 iterations for the PDHH code. We computed 10^6 iterations in the case of the transport

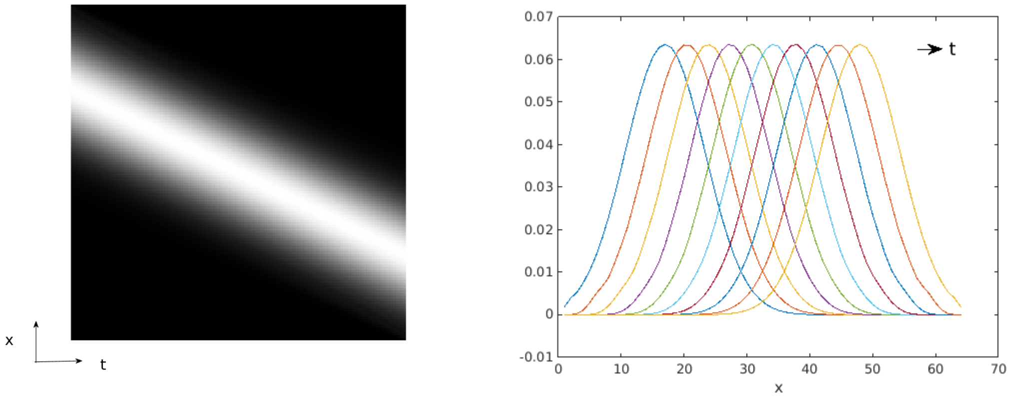


Figure 3: Display of different views of the density $\rho(t, x)$ obtained after 10^6 iterations, for ρ_0 and ρ_1 two Gaussians of the same variance.

between two isotropic Gaussians of the same variance, and we plot the estimated density in Figure 3. The solution, which will be denoted (ρ_s, m_s), is displayed in black and gray, black being 0 and white being 1, in the left image and displayed at different time values superimposed in the right image. We use a grid of $N = 128$ discretization points for ρ_0 and ρ_1 and $M = 128$ points for the time t.
Figure 4 displays, for the example of Figure 3, the L^2-errors between ρ and ρ_s and between m and m_s, the functional $\mathcal{J}$, and the numerical complexity as function of the grid size N, for 5000 iterations, for the three algorithms: *modified* PDPOP, PDHH and PDHHMS. We choose 5000 iterations because this number allows to reach an accurate solution with all of the algorithms. The curves show that despite PDHH has not the best convergence rate during the first iterations, it still converges as quickly as the PDPOP algorithm, while PDHHMS algorithm needs relatively more iterations than both these algorithms. Indeed, the decrease of the functional in the constraint set has not the same behavior as in the PDPOP algorithm, where one has to project onto the divergence-free constraint space. Figure 4 also displays the computation time with respect to the number $M = N$ of discretization points in one direction. It shows that the complexity of the three algorithms is linear in the number of discretization points M^2, and that the PDHH algorithm is 42% faster than the PDPOP algorithm, and that PDHHMS is 78% faster than the PDPOP algorithm in terms of `cputime`.
We compare in Table 1 the number of iterations and the `cputime` required for the error on ρ to drop under a given error. We observe that PDHH and PDHHMS need more iterations to converge, but since each iteration runs faster, we need less `cputime` to reach the desired errors. The explanation for this better `cputime` is that we don't have to solve a Poisson equation at each iteration in both algorithms. Moreover, the PDHHMS algorithm uses a

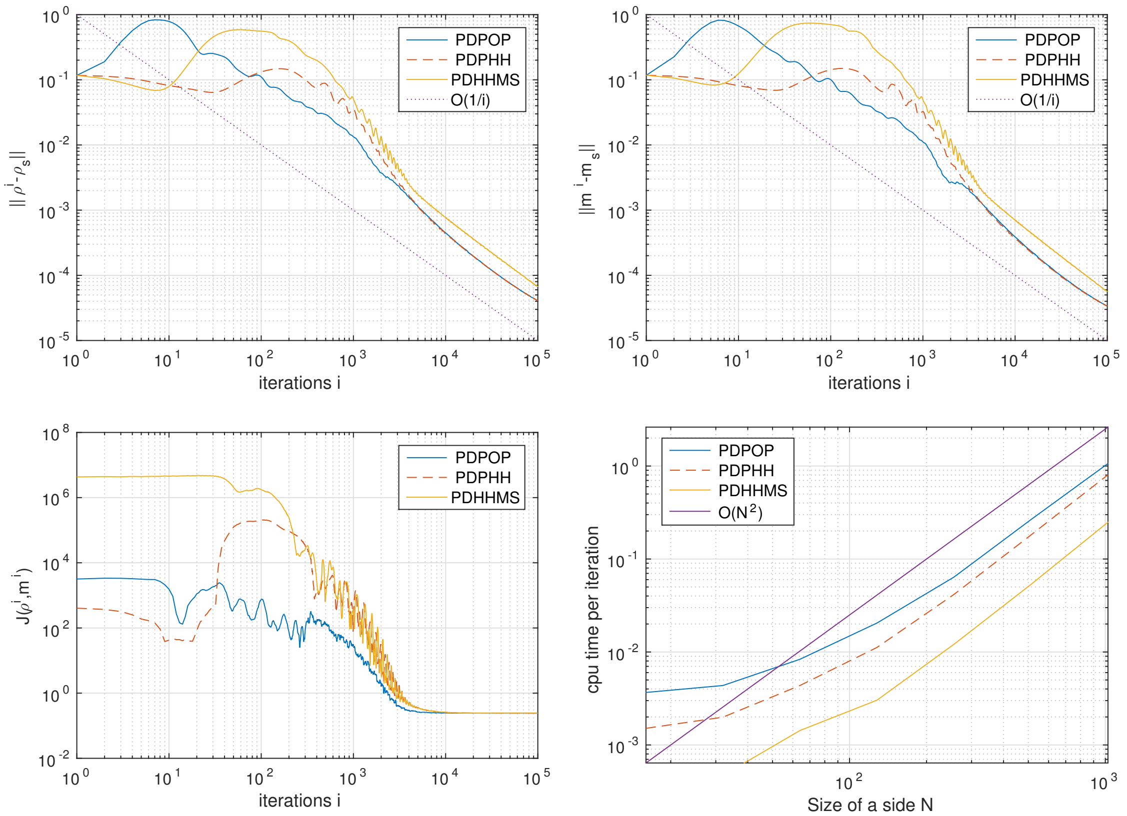


Figure 4: Comparison at each iteration of the L^2-error between ρ and ρ_s (top left) and between m and m_s (top right), the functional $\mathcal{J}$ (bottom left) and the numerical complexity (bottom right), for 5000 iterations, between PDPOP [36], PDHH (algo.3) and PDHHMS (algo.2) algorithms in the case of Figure 3.

simple projection onto a L^2-ball while PDHH uses a projection onto a paraboloid which requires the resolution of a cubic function, as in PDPOP.

Test between an oscillating and a compactly supported density.
Applications of the 1D optimal transport can include analysis of audio signals via Fourier analysis. As an example in 1D we compute the density $\rho(t,.)$ for an oscillating density ρ_1 (Figure 6). The results presented in Figure 5 are obtained for images discretized on a grid 256^2 grid, taking ρ_0 a triangular signal and ρ_1 the absolute value of a sinc function. We observe on Figure 5 that each point of ρ_0 is pushed forward to ρ_1 along a straight line.

4.2.2. Algorithm for 2D+t images. We now consider the transport of two isotropic Gaussians of same variance in two dimensions, and we plot the estimated density in Figure 7: the solution is displayed in black and gray, and will be denoted by (ρ_s, m_s). We use a grid of $N \times P = 64 \times 64$ discretization points for ρ_0 and ρ_1 and $M = 64$ points for the time

$\|\|\rho_i - \rho_s\|\|$	$PDPOP$	$PDHH$	Speedup	$PDHHMS$	Speedup
10^{-2}	1557 (20”)	2241 (15”)	25%	2803 (6”)	67%
10^{-3}	7986 (1’46”)	9408 (1’6”)	38%	12676 (30”)	71%
10^{-4}	56002 (12’36”)	69894 (8’30”)	33%	114295 (4’36”)	63%

Table 1: Performance evaluation in the case of Figure 3. The entries refer to: number of iterations (`cputime`) and the speedup of PDHH (algo.3) and PDHHMS (algo.2) algorithms compared to PDPOP algorithm [36].

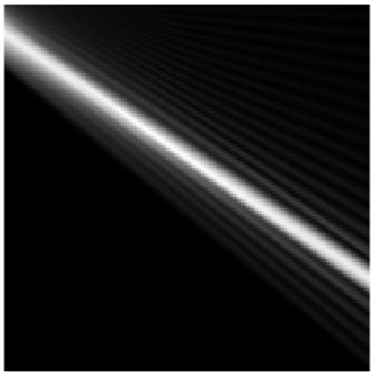
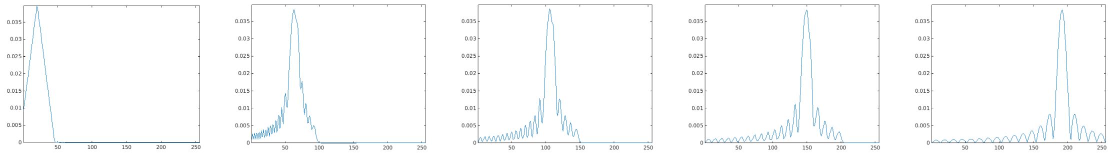


Figure 5: Display of the density $\rho(t,.)$ obtained after 10^6 iterations for ρ_0 a triangular signal and ρ_1 the absolute value of a sinc function.

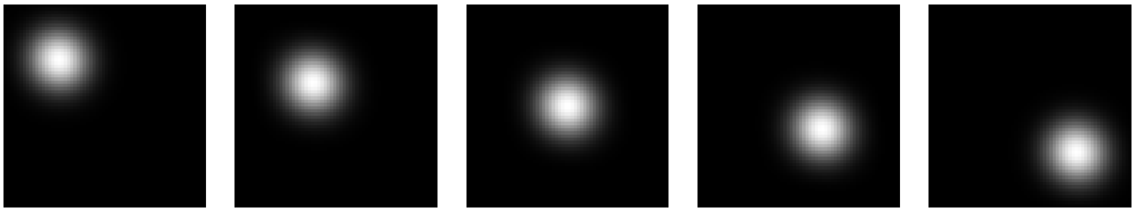
t.

Using $\sigma = 20$, Figure 8 presents the results for the evolution of the functional, and the errors for m and ρ of the example of Figure 7. We first observe that we obtain the $O(1/i)$ convergence rate proved by Chambolle and Pock [12] for the errors on m and ρ, and also that the algorithm has the same behavior than in 1D. Figure 8 shows that even if our algorithm has not the best convergence rate at the beginning, it still converges really quickly until we obtain the $O(1/i)$ convergence rate of the algorithm. Having to remain in the constraint set, the decreasing along the functional is not the same as if we had to project. Figure 8 also displays the computational time with respect to the number of discretization points on a side ($M = N = P$), which shows that the numerical complexity of the two algorithms is $O(M^3)$, *i.e.* linear in the number of discretization points. Beside it shows that, on average, PDHH is 28% faster than PDPOP algorithm. This can be explained by the fact that we don’t have to solve a 3D Poisson equation at each iteration. But unlike PDPOP, we have to evaluate a curl operator, which is slightly time consuming. We compare in Table 2 the number of iterations and the `cputime` required for the error on ρ



t = 0   t = 0.25   t = 0.5   t = 0.75   t = 1

Figure 6: Display of cross sections of the density $\rho(t, x)$ obtained after 10^6 iterations for ρ_0 a triangular signal ($t = 0$) and ρ_1 the absolute value of a sinc function (t=1).



t = 0   t = 0.25   t = 0.5   t = 0.75   t = 1

Figure 7: Display of the density $\rho(t, .)$ obtained after 10^6 iterations.

to drop under a given error for PDHH, PDPOP (the code we implemented) and PDPOP^{gh} which we took on the github page of Peyre [3]. There are two methods to compute the proximal operator of $\mathscr{J}^*$. The first one calculates the proximal operator of $\mathscr{J}$ and then uses the Moreau identity to obtain the proximal operator of $\mathscr{J}^*$, which is the method used in PDPOP^{gh}. The second method calculates directly the proximal operator of $\mathscr{J}^*$, which is done in PDHH and is less time consuming. Therefore we implemented PDPOP (like PDHH) with the second method. Then, (*modified*) PDPOP and PDHH only differs because of the Helmholtz-Hodge decomposition $(\rho, m) = \nabla \times \phi + \nabla h$, and we can really compare its influence on the method.
We observe that PDHH algorithm needs more iterations to converge, but is faster to run one iteration, so it needs less `cputime` to reach the expected error, while PDPOP and PDPOP^{gh} algorithms need the same number of iterations. Indeed, these last two algorithms use equivalent projections. In PDPOP we first test if (ρ, m) is already in the paraboloid and we use the proximal operator of J^* while in PDPOP^{gh} the proximal operator of J is used without test before projecting. Thanks to this new projection and the use of the Helmholtz-Hodge decomposition, an iteration of PDHH is on average 51% faster than an iteration of PDPOP^{gh}.

Test on non convex densities.. The next example of transport considers the case of irregular, non convex and non connected densities with compact supports. Figure 9 shows

[3] https://github.com/gpeyre/2013-SIIMS-ot-splitting

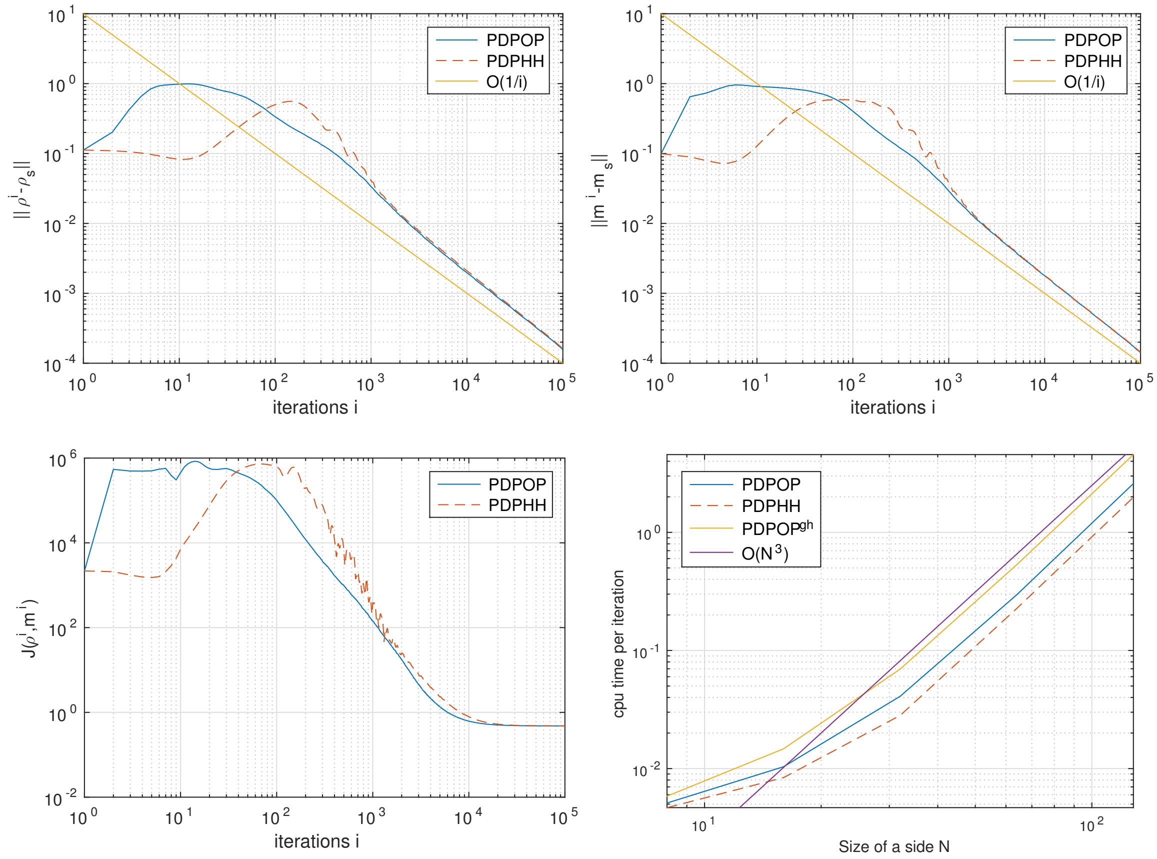


Figure 8: Comparison at each iteration of the functional $\mathcal{J}$ (bottom left), the minimum value of ρ (bottom right), the L^2-errors between ρ and ρ_s (top left) and between m and m_s (top right) between PDPOP (modified algorithm of [36]), PDHH algorithm (algo.3), and also PDPOP^{gh} for the last figure (algo [36] on github), in the case of Figure 7.

the ability of our method to estimate the density $\rho(t,.)$ for such initial and final densities.

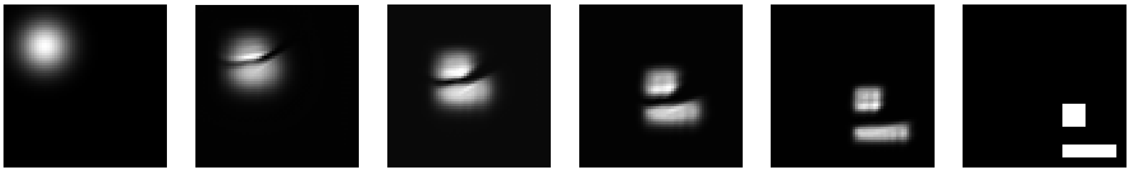
Test on real life images. As last example we compute the density $\rho(t,.)$ for images representing clouds in different positions. The results presented in Figure 10 are obtained for images discretized on a grid $M = 68$ for the time dimension and $N \times P = 120 \times 68$ for the space dimension.

5. Implementation in C++. The implementation in Matlab is slow and can not be used for large images. To overcome this problem we implemented the PDHH algorithm for 2D images in C++ using the PDPOP code developed by Nicolas Boneel, parallelized in OpenMP, that can be downloaded at http://liris.cnrs.fr/~nbonneel/FastTransport/.

If we consider again the table 2 and the real (`wall time`) execution time of Matlab and C++ implementations for the algorithm PDHH, we obtain:

$\|\|\rho_i - \rho_s\|\|$	$PDHH$	$PDPOP$	Speedup	PDPOP^{gh}	Speedup
10^{-2}	1250 (3'15")	1140 (3'58")	18%	1140 (7'10")	55%
10^{-3}	7763 (21'27")	7382 (26'07")	18%	7382 (45'48")	53%
10^{-4}	62616 (3:08'12")	60860 (3:35'55")	13%	60860 (6:18'18")	50%

Table 2: Performance evaluation in the case of Figure 7. The entries refer to: number of iterations (`cputime`) and the speedup of PDHH algorithm (algo. 3) compared to *modified* PDPOP and PDPOP^{gh} algorithms [36].



t = 0 t = 0.2 t = 0.4 t = 0.6 t = 0.8 t = 1

Figure 9: Display of the density $\rho(t,.)$ obtained after 10^6 iterations of a non-convex, non connected density with compact support.

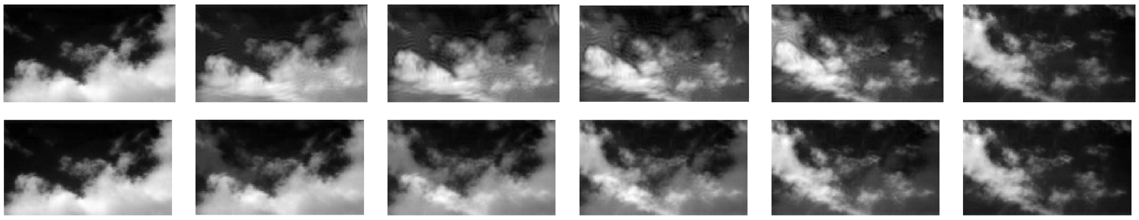
$\|\|\rho^i - \rho^s\|\|$	Matlab	C++	Speedup
10^{-2}	2'37"	13"	$\times 12$
10^{-3}	8'10"	40"	$\times 12$
10^{-4}	1:27'14"	5'44"	$\times 15$

Table 3: Comparison of the execution time of the C++ and Matlab codes in the case of algorithm PDHH.

6. Conclusion. In this article, we introduced a new algorithm for the dynamic optimal transport problem between 1D or 2D images, which respects the divergence-free constraint throughout the iterations, and therefore gets rid of solving a 3D Poisson equation at each iteration in the case of 2D images. We also proved some convexity properties, on the constraint set, of the functional used in this formulation. Besides, this algorithm is easy to implement, faster than state of the art methods for this kind of formulation, and efficient for real-sized images, thanks to a parallelized implementation in C++. Moreover we explained that in 1D+time, it is equivalent to the resolution of a minimal surface equation. Further improvements of the method will include other divergence-free decomposition [27], or other formulations of the primal-dual algorithm.

7. Acknowledgment. We would like to thank Nicolas Papadakis for his kind help. This work was funded by the French Agence Nationale de la Recherche (ANR, Project TOMMI) under reference ANR-11-BS01-014-01.

Appendix A. Calculation of the proximal operator of $||. - \nabla \times h||_1$. In this appendix we compute the proximal operator of $\mathcal{J}_{ms}(x) = ||x - \nabla \times h||_1$. On the discrete grid G^{c1}



t = 0 t = 0.2 t = 0.4 t = 0.6 t = 0.8 t = 1

Figure 10: Display of the density $\rho(t,.)$ of an image of clouds. The first line represents $\rho(t,.)$ obtained after 10^6 iterations of PDHH algorithm while the second line represents the L^2 -interpolation.

defined in 4.1, $||.||_1$ is written

$$||x||_1 = \sum_{k \in G^{c1}} |x_k| \text{ where } |x_k| = \sqrt{(x_k^1)^2 + (x_k^2)^2},$$

and we will denote $F(x) = |x - \nabla \times h|$. Then we have for $\sigma > 0$,

$$\mathrm{prox}_{\sigma \mathcal{J}_{ms}}(x) = \mathrm{prox}_{\sigma ||.-\nabla\times h||_1}(x) = (\mathrm{prox}_{\sigma |.-\nabla \times h|}\, x_k)_{k \in G^{c1}}.$$

We have to compute

$$y' = \mathrm{prox}_F(x) = \mathrm{argmin}_y \frac{1}{2}|y - x|^2 + F(y) = \mathrm{argmin}_y \frac{1}{2}|y - x|^2 + |y - \nabla \times h|. \tag{27}$$

We know that the proximal operator of the ℓ^1-norm $N(x) = |x|$ is given by:

$$\mathrm{prox}_N(x) = \mathrm{argmin}_y \frac{1}{2}|y - x|^2 + |y| = x \ \max(0, \frac{1}{|x|})$$

Since $F(x) = N(x - \nabla \times h)$ one obtains:

$$\begin{aligned} \mathrm{prox}_F(x) &= \nabla \times h + \mathrm{prox}_N(x - \nabla \times h) \\ &= \nabla \times h + (x - \nabla \times h) \ \max(0, \frac{1}{|x - \nabla \times h|}) \end{aligned} \tag{28}$$

For $\sigma > 0$

$$\begin{aligned} \mathrm{prox}_{\sigma F}(x) &= \underset{y}{\mathrm{argmin}} |x - y|^2 + \sigma |y - \nabla \times h|, \\ &= \nabla \times h + (x - \nabla \times h) \ \max(0, 1 - \frac{\sigma}{|x - \nabla \times h|}). \end{aligned}$$

Thanks to Moreau's identity [35] we then obtain

$$
\begin{aligned}
\operatorname{prox}_{\gamma F^*}(x) &= x - \gamma \operatorname{prox}_{F/\gamma}(x/\gamma) \\
&= x - \gamma \left(\nabla \times h + (x/\gamma - \nabla \times h) \max(0, 1 - \frac{1}{\gamma |x/\gamma - \nabla \times h|}) \right) \\
&= x - \gamma \nabla \times h - (x - \gamma \nabla \times h) \max(0, 1 - \frac{1}{|x - \gamma \nabla \times h|}) \\
&= x - \gamma \nabla \times h + (x - \gamma \nabla \times h) \min(0, \frac{1}{|x - \gamma \nabla \times h|} - 1) \\
&= \min(x - \gamma \nabla \times h, \frac{(x - \gamma \nabla \times h)}{|x - \gamma \nabla \times h|})
\end{aligned}
$$

REFERENCES

[1] J. Altschuler, J. Weed, and P. Rigollet, *Near-linear time approximation algorithms for optimal transport via sinkhorn iteration*, in Advances in Neural Information Processing Systems, 2017, pp. 1964–1974.

[2] L. Ambrosio, L. Caffarelli, M. Crandall, and N. Evans, L.C.and Fusco, *Calculus of Variations and Nonlinear Partial Differential Equations*, Springer, 2008.

[3] K. J. Arrow, L. Hurwicz, H. Uzawa, and H. B. Chenery, *Studies in linear and non-linear programming*, vol. 2, Stanford University Press Stanford, 1958.

[4] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2011.

[5] J.-D. Benamou, *Numerical resolution of an unbalanced mass transport problem*, ESAIM: Mathematical Modelling and Numerical Analysis, 37 (2003), pp. 851–868.

[6] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the monge-kantorovich mass transfer problem*, Numerische Mathematik, 84 (2000), pp. 375–393.

[7] A. Bouharguane, E. Maitre, E. Oudet, and N. Papadakis, *Multiphysics optimal transportation and image analysis*, hal preprint : hal-00740671, (2012).

[8] C. Cancès and C. Guichard, *Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure*, Foundations of Computational Mathematics, 17 (2017), pp. 1525–1584.

[9] G. Carlier and C. Poon, *On the total variation wasserstein gradient flow and the tv-jko scheme*, arXiv preprint arXiv:1703.00243, (2017).

[10] J. A. Carrillo and J. S. Moll, *Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms*, SIAM J. Sci. Comput., 31 (2009), pp. 4305–4329.

[11] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical imaging and vision, 20 (2004), pp. 89–97.

[12] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145.

[13] P.-A. Chiappori, R. J. McCann, and L. P. Nesheim, *Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness*, Economic Theory, 42 (2010), pp. 317–354.

[14] L. Chizat, *Unbalanced Optimal Transport: Models, Numerical Methods, Applications*, PhD thesis, PSL Research University, 2017.

[15] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *An interpolating distance between optimal transport and fisher–rao metrics*, Foundations of Computational Mathematics, 18 (2018), pp. 1–44.

[16] P. L. Combettes, L. Condat, J. C. Pesquet, and B. C. Vu, *A forward-backward view of some primal-dual optimization methods in image recovery*, in Image Processing (ICIP), 2014 21st IEEE

International Conference on, IEEE, 2014, pp. 4141–4145.
[17] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-point algorithms for inverse problems in science and engineering, Springer, 2011, pp. 185–212.
[18] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling & Simulation, 4 (2005), pp. 1168–1200.
[19] B. Desjardins, *A few remarks on ordinary differential equations*, Communications in Partial Differential Equations, 21 (1996).
[20] J. H. Fitschen, F. Laus, and G. Steidl, *Dynamic optimal transport with mixed boundary condition for color image processing*, in 2015 International Conference on Sampling Theory and Applications (SampTA), IEEE, 2015, pp. 558–562.
[21] J. H. Fitschen, F. Laus, and G. Steidl, *Transport between rgb images motivated by dynamic optimal transport*, Journal of Mathematical Imaging and Vision, 56 (2016), pp. 409–429.
[22] A. Galichon, *Optimal Transport Methods in Economics*, Princeton University Press, 2016.
[23] V. Girault and P.-A. Raviart, *Finite element methods for Navier-Stokes equations: theory and algorithms*, Springer-Verlag, 1986.
[24] E. Haber, T. Rehman, and A. Tannenbaum, *An efficient numerical method for the solution of the l_2 optimal mass transfer problem*, SIAM Journal on Scientific Computing, 32 (2010), pp. 197–211.
[25] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for total variation image restoration*, tech. report, 2790, Optimization Online, 2010.
[26] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 119–149.
[27] M. Henry, *Transport optimal et ondelettes : nouveaux algorithmes et applications à l'image*, PhD thesis, Grenoble Alpes University, 2016.
[28] M. Henry, E. Maitre, , and V. Perrier, *Optimal transport using helmholtz-hodge decomposition and first order primal-dual algorithm*, in 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 4748–4752.
[29] R. Hug, *Analyse mathématique et convergence d'un algorithme pour le transport optimal dynamique : cas des plans de transports non réguliers, ou soumis à des contraintes*, PhD thesis, Grenoble Alpes University, 2016.
[30] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the fokker–planck equation*, SIAM journal on mathematical analysis, 29 (1998), pp. 1–17.
[31] S. Kadri Harouna and V. Perrier, *Divergence-free wavelet projection method for incompressible viscous flow on the square*, Multiscale Modeling & Simulation, 13 (2015), pp. 399–422.
[32] S. Kolouri, S. Park, M. Thorpe, D. Slepcev, , and G. Rohde, *Optimal mass transport: Signal processing and machine-learning applications*, IEEE Signal Process Mag., 34 (2017), pp. 43–59.
[33] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, *Optimal mass transport: Signal processing and machine-learning applications*, IEEE signal processing magazine, 34 (2017), pp. 43–59.
[34] D. Lombardi and E. Maitre, *Eulerian models and algorithms for unbalanced optimal transport*, ESAIM: Mathematical Modelling and Numerical Analysis, 49 (2015), pp. 1717–1744.
[35] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bulletin de la Société mathématique de France, 93 (1965), pp. 273–299.
[36] N. Papadakis, G. Peyré, and E. Oudet, *Optimal transport with proximal splitting*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 212–238.
[37] G. Peyré and M. Cuturi, *Computational Optimal Transport*, vol. 11, Foundations and Trends in Machine Learning, 2019. arXiv:1803.00567v2.
[38] G. Peyré, J. Fadili, and J. Rabin, *Wasserstein active contours*, in Image Processing (ICIP), 2012 19th IEEE International Conference on, IEEE, 2012, pp. 2541–2544.
[39] B. Piccoli and F. Rossi, *Generalized wasserstein distance and its application to transport equations with source*, Archive for Rational Mechanics and Analysis, 211 (2014), pp. 335–358.
[40] J. Rabin, S. Ferradans, and N. Papadakis, *Adaptive color transfer with relaxed optimal transport*, in Image Processing (ICIP), 2014 21st IEEE International Conference on, IEEE, 2014, pp. 4852–4856.

[41] R. T. Rockafellar, *Convex analysis*, vol. 28, Princeton university press, 1997.
[42] Y. Rubner, C. Tomasi, and L. J. Guibas, *A metric for distributions with applications to image databases*, in Computer Vision, 1998. Sixth International Conference on, IEEE, 1998, pp. 59–66.
[43] M. A. Schmitz, M. Heitz, N. Bonneel, F. Ngole, D. Coeurjolly, M. Cuturi, G. Peyré, and J.-L. Starck, *Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning*, SIAM Journal on Imaging Sciences, 11 (2018), pp. 643–678.
[44] T. Valkonen, *A primal-dual hybrid gradient method for non-linear operators with applications to MRI*, Inverse Problems, 30 (2014), p. :055012.
[45] C. Villani, *Topics in optimal transportation*, no. 58, American Mathematical Soc., 2003.
[46] R. Yildizoglu, J.-F. Aujol, and N. Papadakis, *Active contours without level sets*, in Image Processing (ICIP), 2012 19th IEEE International Conference on, IEEE, 2012, pp. 2549–2552.