

MULTI-PHYSICS OPTIMAL TRANSPORTATION AND IMAGE INTERPOLATION

ROMAIN HUG¹, EMMANUEL MAITRE¹ AND NICOLAS PAPADAKIS²

Abstract. Optimal transportation theory is a powerful tool to deal with image interpolation. This was first investigated by [Benamou and Brenier, *Numer. Math.* **84** (2000) 375–393.] where an algorithm based on the minimization of a kinetic energy under a conservation of mass constraint was devised. By structure, this algorithm does not preserve image regions along the optimal interpolation path, and it is actually not very difficult to exhibit test cases where the algorithm produces a path of images where high density regions split at the beginning before merging back at its end. However, in some applications to image interpolation this behaviour is not physically realistic. Hence, this paper aims at studying how some physics can be added to the optimal transportation theory, how to construct algorithms to compute solutions to the corresponding optimization problems and how to apply the proposed methods to image interpolation.

Mathematics Subject Classification. 68U10, 65K10, 35D05.

Received April 23, 2015. Revised April 29, 2015.
Published online November 5, 2015.

1. INTRODUCTION

While optimal transportation theory has been widely studied in the last years [34], the numerical resolution of the problem is still a challenge. In the seminal work of Benamou and Brenier [3], instead of computing directly the transport, the output of their algorithm was the geodesic path between the two densities according to the Wasserstein metric. This is of main interest for image interpolation purposes and it has been widely used in the image processing community for registration [23], optical flow estimation [16], deblurring of image sequences in microscopy [12], *etc.* In the original approach, the authors of [3] demonstrates that the optimal transportation problem can be obtained through the minimization of the kinetic energy using flows which preserve the mass. Using a convex reformulation of the problem, they introduced an algorithm that computes the geodesic path between given images, by using a Augmented Lagrangian method [21]. As the mass transfer associated to a L^2 ground cost follows straight lines [11], such property has been considered in order to speed-up computations [26]. Approximation with Knothe transport [7, 13] or linearization of the problem for close densities [10] have also been considered to propose faster algorithms.

Keywords and phrases. Optimal transportation, image multiphysics, proximal splitting method, non-convex optimization.

¹ Laboratoire Jean Kuntzmann, Grenoble University, Université Joseph Fourier and CNRS, France.
romain.hug@imag.fr, emmanuel.maitre@imag.fr

² Institut de Mathématiques de Bordeaux, CNRS and Université Bordeaux 1, France.
nicolas.papadakis@math.u-bordeaux.fr

Structurally, optimal transportation does not preserve image regions along the optimal interpolation path. Consequently it is actually not difficult to exhibit test cases where these algorithms produce a path of images where high density regions split at the beginning before re-merging (see [3] for an example). However, for image interpolation applications, this behavior is not desirable. Suppose for instance that the images represent the evolution in time of a droplet of oil in water. The droplet can obviously split during time; however, this phenomenon has a physical cost which is not taken into account in the interpolation given by the genuine optimal transportation model. For an elastic membrane we would expect no splitting at all!

Hence, this article aims at studying of how physics can be accounted for in the optimal transportation theory. We study the anisotropic optimal transportation in order to model domains that could contain obstacles, polarization, field of force, *etc.* We consider physical priors on the velocities that transport the mass, in order to prevent the mass transfer to follow straight lines and avoid mass splitting. To that end, we introduce several extended models of optimal transportation which will, on one hand, add other energy terms to the classical kinetic energy, or, on the other hand, add constraints on the admissible space for minimizers. We will namely focus on the convex transport on Riemannian manifolds and the non convex transports induced by physical priors. We also describe how to construct first-order algorithms to compute critical points to the corresponding optimization problems and how to apply the proposed methods to image interpolation. Notice that due to this generalized form, the associated Euler–Lagrange equations are not of Monge–Ampère type and second-order algorithms based on this formulation [29], and recently [4–6, 18], can not be used in our context. The dynamical optimal transportation formulation is also the only one that can handle vanishing densities (without any assumption on the support of the densities), which is also of interest for image processing applications.

In a previous work [9], we already considered physical priors but the optimization tools were not satisfying in practice. Following [30], we here propose to use proximal splitting methods to optimize properly the defined problems. With respect to these two papers, our contributions are the following. We study the well-posedness of anisotropic optimal transportation and generalized optimal transportations that contain regularizations of the velocity. We provide adequate numerical schemes to solve the defined problems. We apply our model in order to interpolate high-resolution oceanographic images.

This paper is organized as follows. We first review in Section 2 the numerical scheme introduced in [30] in order to solve the convexified optimal problem of [3]. Next, we detail how such framework can be extended to deal with anisotropic optimal transportation in Section 3. Hence, we introduce in Section 4 some non-convex functionals that involve physical priors on the velocity that transfer the mass. Numerical validations of Section 5 demonstrate the interest of our algorithms on image interpolation examples.

2. DYNAMICAL OPTIMAL TRANSPORTATION FORMULATION

2.1. Optimal transportation

In the following, we restrict our exposition to maps $T : \Omega \subset \mathbb{R}^d \mapsto \Omega$ where $d > 0$ is the dimension of the problem and Ω is a convex and bounded domain of \mathbb{R}^d . Let ρ_0, ρ_1 be two nonnegative L^1 functions on Ω , of equal mass: we will assume without loss of generality that

$$\int_{\Omega} \rho_0(\mathbf{x}) \, d\mathbf{x} = \int_{\Omega} \rho_1(\mathbf{x}) \, d\mathbf{x} = 1. \quad (2.1)$$

A valid transport map T is a map that pushes forward the measure $\rho_0(\mathbf{x}) \, d\mathbf{x}$ onto $\rho_1(\mathbf{x}) \, d\mathbf{x}$. In term of densities, when T is smooth and one-to-one this corresponds to the constraint

$$\rho_0(\mathbf{x}) = \rho_1(T(\mathbf{x})) |\det(\nabla T(\mathbf{x}))|$$

where $\nabla T(\mathbf{x}) \in \mathbb{R}^{d \times d}$ is the differential of T at \mathbf{x} . This is known as the Jacobian’s equation. We call $\mathcal{T}(\rho_0, \rho_1)$ the set of transport maps that satisfy this constraint. An optimal transport T solves

$$\min_{T \in \mathcal{T}(\rho_0, \rho_1)} \int C(\mathbf{x}, T(\mathbf{x})) \, d\mathbf{x} \quad (2.2)$$

where $C(\mathbf{x}_0, \mathbf{x}_1) \geq 0$ is the cost of assigning $\mathbf{x}_0 \in \Omega$ to $\mathbf{x}_1 \in \Omega$. In the case $C(\mathbf{x}_0, \mathbf{x}_1) = \|\mathbf{x}_0 - \mathbf{x}_1\|^2$, the optimal value of (2.2), the so-called optimal transport distance, is often called the L^2 -Wasserstein distance between the densities ρ_0 and ρ_1 .

2.2. Fluid mechanics formulation

The geodesic path between the measures with densities ρ_0 and ρ_1 can be shown to have density $t \mapsto \rho(t, \mathbf{x})$ where $t \in [0, 1]$ parameterizes the path and

$$\rho_0(\mathbf{x}) = \rho(t, T_t(\mathbf{x})) |\det(\nabla T_t(\mathbf{x}))| \text{ where } T_t = (1 - t) \text{Id}_d + tT.$$

Benamou and Brenier showed in [3] that this geodesic solves the following non-convex problem over the densities $\rho(t, \mathbf{x}) \in \mathbb{R}_+$ and a velocity field $v(t, \mathbf{x}) \in \mathbb{R}^d$

$$\min_{(\rho, v) \in C^0} \frac{1}{2} \int_0^1 \int_{\Omega} \rho(t, \mathbf{x}) \|v(t, \mathbf{x})\|^2 \, d\mathbf{x} \, dt, \tag{2.3}$$

under the set of non-linear constraints

$$C^0 = \{(\rho, v), \partial_t \rho + \text{div}_{\mathbf{x}}(\rho v) = 0, \langle v, \mathbf{n} \rangle = 0, \rho(0, \cdot) = \rho_0, \rho(1, \cdot) = \rho_1\}, \tag{2.4}$$

where the first relation in C^0 is the continuity equation. We impose homogeneous Neumann conditions on the velocity field v , through \mathbf{n} which is the normal of the domain Ω .

Following [3] and introducing the change of variable $(\rho, v) \mapsto (\rho, m)$, where m is the momentum $m = \rho v$, we obtain a convex optimization problem over the couple (ρ, m) :

$$\min_{(\rho, m) \in C} \mathcal{J}(\rho, m) := \int_0^1 \int_{\Omega} J(\rho, m) \, d\mathbf{x} \, dt$$

with

$$J(\rho, m) = \begin{cases} \frac{\|m\|^2}{2\rho} & \text{if } \rho > 0 \\ 0 & \text{if } (\rho, m) = (0, 0) \\ +\infty & \text{otherwise.} \end{cases} \tag{2.5}$$

The set of constraint C becomes linear and reads:

$$C(\rho, m) = \{(\rho, m), \partial_t \rho + \text{div}_{\mathbf{x}} m = 0, \langle m, \mathbf{n} \rangle = 0, \rho(0, \cdot) = \rho_0, \rho(1, \cdot) = \rho_1\}. \tag{2.6}$$

Denoting as $\iota_C(x) = 0$ if $x \in C$, $+\infty$ otherwise, the convex model of Benamou–Brenier [3] becomes:

$$(\rho^*, m^*) = \underset{\rho, m}{\text{argmin}} \mathcal{J}(\rho, m) + \iota_C(\rho, m). \tag{2.7}$$

The existence of minimizers of such problem in the space of measures has been studied in [34]. The case of L^2 minimizers is the subject of an ongoing work [25].

2.3. Discretization

We here recall the discretization proposed in [30]. For simplicity of exposure, we describe the discretization for the 1-D case and $\Omega = [0; 1]$. It extends verbatim to higher dimensional discretization $d > 1$.

2.3.1. Centered grid

We denote $P + 1$ the number of discretization points in time, and $N + 1$ the number of discretization points in space. We introduce the centered grid discretizing the space-time square $[0, 1]^2$ in $(P + 1) \times (N + 1)$ points as

$$\mathcal{G} = \{(t_k = k/P, \mathbf{x}_i = i/N) \in [0, 1]^2, 0 \leq k \leq P, 0 \leq i \leq N\}.$$

We denote

$$x = (\bar{\rho}, \bar{m}) \in \mathcal{E}_c = (\bar{\rho}_{k,i}, \bar{m}_{k,i})_{\substack{0 \leq k \leq P \\ 0 \leq i \leq N}}$$

the variables discretized on the centered grid, where $\mathcal{E}_c = (\mathbb{R}^{d+1})^{\mathcal{G}} = (\mathbb{R}^2)^{\mathcal{G}}$ is the finite dimensional space of centered variables.

2.3.2. Staggered grid

The use of a staggered grid is very natural in the context of the discretization of a divergence operator associated to a vector field $y = (y^i)_{i=1}^d$ on \mathbb{R}^d (we here focus on the case $d = 2$, the first dimension being temporal and the second one spatial). The basic idea is to perform an accurate evaluation of every partial derivative $\partial_{\mathbf{x}_i} y^i$ at prescribed nodes of a cartesian grid using standard centered finite differences. One way to perform this computation is to impose to the grid on which the y^i scalar field is defined to be centered on $N_i + 1$ points along the \mathbf{x}_i direction, while a center grid would have N_i points on this dimension. This simple requirement forces the y^i scalar fields to be defined on different grids for each dimension i . The resulting discrete vector field gives us the possibility to evaluate the divergence operator by a uniform standard centered scheme which is not possible using a single grid of discretization for every component y^i . As a consequence, similarly to the discretization of PDE's in incompressible fluid dynamics (see for instance [24]), we consider a staggered grid discretization which is more relevant to deal with the continuity equation, and is defined as

$$\begin{aligned} \mathcal{G}^t &= \left\{ (t_k = (k + 1/2), \mathbf{x}_i = i/N)/P \in \frac{[-1, 2P + 1]}{2P} \times [0, 1], -1 \leq k \leq P, 0 \leq i \leq N \right\}, \\ \mathcal{G}^x &= \left\{ (t_k = k/P, \mathbf{x}_i = (i + 1/2)/N) \in [0, 1] \times \frac{[-1, 2N + 1]}{2N}, 0 \leq k \leq P, -1 \leq i \leq N \right\}. \end{aligned}$$

From these definitions, we see that \mathcal{G}^t contains $(P+2) \times (N+1)$ points and \mathcal{G}^x corresponds to a $(P+1) \times (N+2)$ discretization. We finally denote

$$y = (\rho, m) \in \mathcal{E}_s = ((\rho_{k,i})_{\substack{-1 \leq k \leq P \\ 0 \leq i \leq N}}, (m_{k,i})_{\substack{0 \leq k \leq P \\ -1 \leq i \leq N}})$$

the variables discretized on the staggered grid, where $\mathcal{E}_s = \mathbb{R}^{\mathcal{G}^t} \times \mathbb{R}^{\mathcal{G}^x}$ is the finite dimensional space of staggered variables.

2.3.3. Interpolation and divergence operators

We introduce a midpoint interpolation operator $\mathcal{I} : \mathcal{E}_s \rightarrow \mathcal{E}_c$, where, for $y = (\rho, m) \in \mathcal{E}_s$, we define $\mathcal{I}(y) = x = (\bar{\rho}, \bar{m}) \in \mathcal{E}_c$ as

$$\forall 0 \leq k \leq P, \quad \forall 0 \leq i \leq N, \quad \begin{cases} \bar{\rho}_{k,i} = (\rho_{k-1,i} + \rho_{k,i})/2, \\ \bar{m}_{k,i} = (m_{k,i-1} + m_{k,i})/2. \end{cases} \quad (2.8)$$

The space-time divergence operator $\text{div}_{t,\mathbf{x}} : \mathcal{E}_s \rightarrow \mathbb{R}^{\mathcal{G}}$ is defined, for $y = (\rho, m) \in \mathcal{E}_s$ as

$$\forall 0 \leq k \leq P, \quad \forall 0 \leq i \leq N, \quad \text{div}_{t,\mathbf{x}}(y)_{k,i} = (\rho_{k,i} - \rho_{k-1,i}) + (m_{k,i} - m_{k,i-1}).$$

2.3.4. Final convex problem

With the previously introduced grids, the discrete problem 2.7 can be reformulated as:

$$\begin{aligned}
 (\rho^*, m^*) &= \operatorname{argmin}_{(\rho, m) \in \mathcal{E}_s} \mathcal{J}(\mathcal{I}(\rho, m)) + \iota_C(\rho, m), \\
 &= \operatorname{argmin}_{(\rho, m) \in \mathcal{E}_s} \sum_{t=0}^1 \sum_{\mathbf{x} \in \Omega} J(\mathcal{I}(\rho(t, \mathbf{x}), m(t, \mathbf{x}))) + \iota_C(\rho, m).
 \end{aligned}
 \tag{2.9}$$

The problem (2.9) consists in minimizing the sum of two non smooth convex functionals. As in [30], we can make use of proximal splitting methods to solve the problem.

2.4. Optimization with proximal splitting

In [30], the convex problem 2.9 is optimized using the primal dual algorithm of [15]. Such algorithm is dedicated to the optimization of problems of the form $G_1(Ku) + G_2(u)$, where the G_i 's functions are convex and K is a linear operator ($K = \mathcal{I}$ for the formulation of optimal transportation with staggered and regular grids). In this paper, as we will consider more general problems, we instead rely on the Generalized Forward–Backward (GFB) algorithm proposed in [31]. This algorithm is dedicated to problems of the form:

$$\operatorname{argmin}_u \sum_{i=1}^n G_i(u) + F(u),
 \tag{2.10}$$

where the additional term F is a differentiable convex function with a L -Lipschitz-continuous gradient ∇F and G_i 's are assumed proper, convex and simple. By simple, we mean that the proximal operator of the functions G_i , defined as

$$\operatorname{Prox}_{\tau G}(u) = \operatorname{argmin}_z \frac{\|z - u\|^2}{2\tau} + G(z)$$

can be written in closed form. The GFB algorithm reads:

Algorithm 1. Generalized Forward-Backward.

```

Set  $N \in \mathbb{N}^+$ ,  $\epsilon > 0$ ,  $\tau \in ]0; 1/L[$ ,  $\lambda \in ]0; 1]$  and  $\sum_{i=1}^n \omega_i = 1$ ,  $\omega_i \geq 0$  and initialize  $z_i^0 = u^0$ ,  $i = 1 \dots n$ 
for  $k = 0 \dots N$  do
  while  $\|u^k - u^{k-1}\| > \epsilon$  do
     $z_i^{k+1} = z_i^k + \lambda (\operatorname{Prox}_{\tau/\omega_i G_i}(2u^k - z_i^k - \tau \nabla F(u^k)) - u^k)$ ,  $i = 1 \dots n$ ,
     $u^{k+1} = \sum_{i=1}^n \omega_i z_i^{k+1}$ 
  end while
end for

```

If F is uniformly convex, then the GFB strongly converges to a minimizer of (2.10), otherwise, the convergence is weak. This is a generalization of the Forward–Backward (FB) algorithm that deals with problems of the form:

$$\operatorname{argmin}_u G(u) + F(u),$$

where G is a convex proper and simple function and F is convex, differentiable with a L -Lipschitz-continuous gradient. The FB algorithm is presented in Algorithm 2.

Notice that primal dual methods have been recently extended [8, 14, 17] to deal with the problem (2.10).

Algorithm 2. Forward-Backward.

```

Set  $N \in \mathbb{N}^+$ ,  $\epsilon > 0$ ,  $\tau \in ]0; 1/L[$  and initialize  $u^0$ 
for  $k = 0 \dots N$  do
  while  $\|u^k - u^{k-1}\| > \epsilon$  do
     $u^{k+1} = \text{Prox}_{\tau G}(u^k - \tau \nabla F(u^k))$ 
  end while
end for

```

2.5. Application to problem (2.9)

In order to deal with the different numerical grids, we need to define a variable $x = (\bar{\rho}, \bar{m})$ on the regular grid and a variable $y = (\rho, m)$ on the staggered one. The linear set of constraint $S(x, y) = \{\mathcal{I}(y) = x\}$ is therefore required to couple these two variables. Hence, the problem (2.9) can be reformulated as:

$$(x^*, y^*) = \underset{x, y}{\operatorname{argmin}} \mathcal{J}(x) + \iota_C(y) + \iota_S(x, y).$$

The GFB algorithm 1 can finally be applied by taking $F = 0$, $G_1(x, y) = \mathcal{J}(x) + \iota_C(y)$ and $G_2(x, y) = \iota_S(x, y)$. We now detail how computing the proximal operators of the different terms involved in this problem.

2.5.1. Computing $\operatorname{Prox}_{\tau \mathcal{J}}$

Since the functional \mathcal{J} is separable, the proximal operator can be computed independently as the proximal operator of J for each point $(\tilde{\rho}, \tilde{m})(t, \mathbf{x}) \in \mathbb{R} \times \mathbb{R}^d$ with $(t, \mathbf{x}) \in [0; 1] \times \Omega$. Notice that J is a 1-homogeneous functional since $J(k\rho, km) = kJ(\rho, m)$. As a consequence, its Legendre–Fenchel transform is the indicator of a convex set. This corresponds to the projection $q^* = P_K(\tilde{q})$ on the convex set $K \{q = (q_1, q_2, q_3) \text{ s.t. } q_1 + (q_2^2 + q_3^2)/2 \leq 0\}$ in the step B of the algorithm of [3]. One can show [30] that this projection can be equivalently obtained by the pointwise computation of the proximal operator of J at point $(\tilde{\rho}, \tilde{m})$. Denoting as ι_K the indicator of the convex set K ($\iota_K(q) = 0$ if $q \in K$ and $+\infty$ otherwise), the proximal operator of ι_K just computes the projection onto the convex set K so that

$$q^* = \operatorname{Prox}_{\gamma \iota_K}(\tilde{q}) = P_K(\tilde{q}).$$

From Moreau’s identity, we have that:

$$\operatorname{Prox}_{\gamma F^*}(x) = x - \gamma \operatorname{Prox}_{F/\gamma}(x/\gamma).$$

Denoting $\tilde{q} = (\tilde{\rho}, \tilde{m})$ is then possible to show that

$$\operatorname{Prox}_{\iota_K^*}(\tilde{\rho}, \tilde{m}) = (\tilde{\rho}, \tilde{m}) - P_K(\tilde{\rho}, \tilde{m}) = \operatorname{Prox}_J(\tilde{\rho}, \tilde{m}),$$

so that:

$$P_K(\tilde{\rho}, \tilde{m}) = (\tilde{\rho}, \tilde{m}) - \operatorname{Prox}_J(\tilde{\rho}, \tilde{m}),$$

which shows that the projection onto K can be computed through the computation of the proximal operator.

The following proposition presented in [30] shows that the functional J defined in (2.5) is simple, in the sense that its proximal operator can be computed in closed form.

Proposition 2.1. *One has for all $(\tilde{\rho}, \tilde{m}) \in \mathbb{R} \times \mathbb{R}^d$,*

$$\operatorname{Prox}_{\tau J}(\tilde{\rho}, \tilde{m}) = \begin{cases} (\rho^*, \mu(\rho^*)) & \text{if } \rho^* > 0, \\ (0, 0) & \text{otherwise.} \end{cases}$$

where

$$\forall \rho \geq 0, \quad \mu(\rho) = \frac{\rho \tilde{m}}{\rho + \tau}$$

and ρ^* is the largest real root of the third order polynomial equation in ρ

$$P(\rho) = (\rho - \tilde{\rho})(\rho + \tau)^2 - \frac{\tau}{2} \|\tilde{m}\|^2 = 0.$$

2.5.2. Computing Prox_{ι_C}

The proximal mapping of ι_C is Proj_C the orthogonal projector on the convex set C .

We first extract the boundary values on the staggered grid using the linear operator b , defined, for $\tilde{y} = (\tilde{\rho}, \tilde{m})$. For $d = 2$ and $m = (m_1, m_2)$, if we consider $\Omega = [0; 1]^2$ and $t \in [0; 1]$, this operator reads:

$$b(\tilde{y}) = (\tilde{\rho}(0, \cdot, \cdot), \tilde{\rho}(1, \cdot, \cdot), \tilde{m}_1(\cdot, 0, \cdot), \tilde{m}_1(\cdot, 1, \cdot), \tilde{m}_2(\cdot, \cdot, 0), \tilde{m}_2(\cdot, \cdot, 1)).$$

Denoting the boundary conditions as

$$b_0 = (\rho_0, \rho_1, 0, 0, 0, 0).$$

The affine set C that can be rewritten as

$$C = \{y = (\rho, m), Ay = z\}$$

where

$$Ay = (b(y), \text{div}_{t,\mathbf{x}}(y)), \text{ and } z = (b_0, 0).$$

This projection can be computed by solving a linear system as

$$\text{Proj}_C(\tilde{y}) = (\text{Id} - A^* \Delta^{-1} A)\tilde{y} + A^* \Delta^{-1} z$$

where applying $\Delta^{-1} = (AA^*)^{-1}$ requires solving a Poisson equation on the centered grid with the prescribed boundary conditions. This can be achieved with Fast Fourier Transform in $O(NP \log(NP))$ operations where N and P are number of spatial and temporal points, see [32].

2.5.3. Computing Prox_{ι_S}

The proximal mapping of ι_S is Proj_S the orthogonal projector on the convex set $S(x, y) = \{\mathcal{I}(y) = x\}$. It is obtained as $(x, y) = \text{Proj}_S(\tilde{x}, \tilde{y}) = (\mathcal{I}(z), z)$ where

$$z = (\text{Id} + \mathcal{I}^* \mathcal{I})^{-1}(\tilde{y} + \mathcal{I}^*(\tilde{x})), \tag{2.11}$$

and \mathcal{I}^* is the adjoint of the linear interpolation operator. Note that computing Proj_S requires solving a linear system, but this system is separable along each dimension of the discretization grid, so it only requires solving a series of small linear systems. Furthermore, since the corresponding inverse matrix is the same along each dimension, we pre-compute explicitly the inverse of these $d + 1$ matrices.

2.6. Limitations and motivations

When solving the optimal transportation problem between two densities, we know that the mass transfer will follow straight lines. The structures contained in the original data can thus disappear along the optimal path. Such behavior is illustrated in Figure 1 that presents the optimal path between images containing two Gaussians. This result has been obtained by taking $\tau = 0.1$ and $\omega_1 = \omega_2 = \frac{1}{2}$ with the GFB algorithm.

In this work, we aim at incorporating some physical priors in order to preserve the structure contained in the data along the optimal path. This will be of main interest for geoscience imaging applications. Ocean and atmosphere, that are observed with satellites, are indeed driven by complex physical laws. The temporal interpolation of such satellite images is an important problem in this community and it should correspond to the underlying dynamics. As interpolation with classic optimal transportation involves constant mass transport along straight lines, this is not a satisfactory solution. Some generalizations are now proposed to tackle such issues.

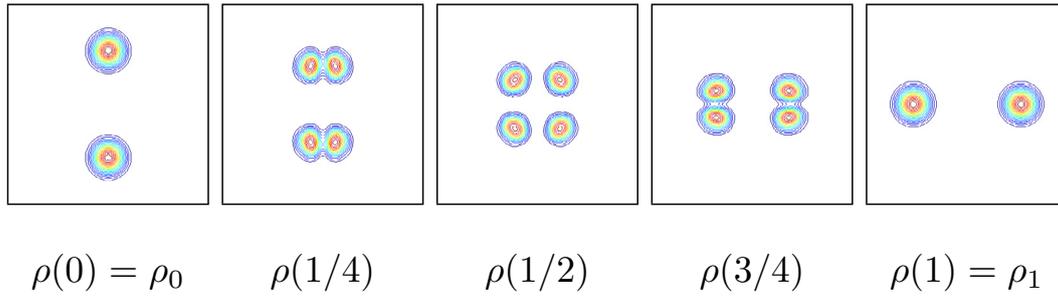


FIGURE 1. Two gaussians experiments: the gaussians split along the optimal path between ρ_0 and ρ_1 .

3. ANISOTROPIC OPTIMAL TRANSPORTATION

We now propose to study the optimal transportation problem within an anisotropic environment that could contain obstacles, polarization, field of force, *etc.* To that end, we search for a continuous transport of mass (ρ, v) which minimizes an energy defined as

$$\int_0^1 \int_{\Omega} \rho(t, \mathbf{x}) v(t, \mathbf{x})^T A(t, \mathbf{x}) v(t, \mathbf{x}) \, d\mathbf{x} \, dt, \tag{3.1}$$

with $A(t, \mathbf{x})$ a symmetric positive definite matrix that represents the anisotropic penalization of the displacement energy. Such modelling is of interest in many applications such as meteorology or oceanography. For instance, if we want to interpolate or measure the distance (in the Wasserstein sense) between two meteorologic images, we should take into account the atmospheric constraints: air mass displacements, rotating motion (cyclones and anticyclones), *etc.* We will show in this section that the same formulations as before can naturally deal with such anisotropic problem.

3.1. Existence

When anisotropy just varies in space (*i.e.* $A(t, \mathbf{x}) = A(\mathbf{x})$), one can get back to a classic Monge–Kantorovitch problem on Riemannian manifold. In the case of a real time dependence, even if the numerical algorithms are still providing a seemingly sound solution, there are no equivalence with any optimal transportation problem and the existence and uniqueness of a solution remain an open subject. As a consequence, we here focus on the case $A(t, \mathbf{x}) = A(\mathbf{x})$. Denoting as M a Riemannian manifold and d the associate geodesic distance, let us first remind McCann’s theorem [27].

Theorem 3.1 (McCann). *Let μ be a probability measure absolutely continuous with respect to the volume measure on M . We assume that μ is also absolutely continuous with respect to all probability measures ν on M which are linked to μ with a transference plan of finite cost. Then there exists a unique optimal transference plan $T(\mathbf{x}) = \exp_{\mathbf{x}}(-\nabla\psi(\mathbf{x}))$ (defined for μ -a.e. x , with $\psi : M \mapsto \mathbb{R}$ a d^2 -concave function) for the Monge transportation problem:*

$$\int d(\mathbf{x}, T(\mathbf{x}))^2 d\mu(\mathbf{x}) = \inf_{S, S\#_{\mu}=\nu} \int d(\mathbf{x}, S(\mathbf{x}))^2 d\mu(\mathbf{x})$$

T is uniquely determined μ -almost everywhere. Moreover, for μ almost all $\mathbf{x} \in M$, the path $\gamma(t) = \exp_{\mathbf{x}}(-t\nabla\psi(\mathbf{x}))$, $t \in [0, 1]$ is the unique minimizing geodesic who joins \mathbf{x} to $T(\mathbf{x})$.

In the following, we will take $M = \Omega \subset \mathbb{R}^2$. We endow Ω with the Riemannian metric g_A which associates to $x \in \Omega$ the scalar product on $\mathbb{R}^2 : (Y, Z) \mapsto Y^T A(\mathbf{x})Z$. Next, we can define the distance on Ω as:

$$\forall (\mathbf{x}, \mathbf{y}) \in \Omega, \, d_A(\mathbf{x}, \mathbf{y}) = \inf_{\gamma} \left\{ \int_0^1 \sqrt{g_A(\dot{\gamma}(t), \dot{\gamma}(t))} \, dt = \int_0^1 (\dot{\gamma}(t)^T A(\gamma(t)) \dot{\gamma}(t))^{\frac{1}{2}} \, dt; \, \gamma(0) = \mathbf{x}, \, \gamma(1) = \mathbf{y} \right\}.$$

The Monge’s problem on the Riemannian manifold (Ω, g_A) for the cost $c = d_A^2$ consists in looking for a transport map T that satisfies:

$$\int_{\Omega} d_A(\mathbf{x}, T(\mathbf{x}))^2 \rho_0(\mathbf{x}) \, d\mathbf{x} = \inf_{S \# \rho_0 = \rho_1} \int_{\Omega} d_A(\mathbf{x}, S(\mathbf{x}))^2 \rho_0(\mathbf{x}) \, d\mathbf{x}.$$

The following proposition shows that this Monge’s problem can be solved by finding $(\rho^*, v^*) \in C^0$, the set of mass preserving constraints defined in relation (2.4), that minimizes the transport cost (3.1) with $A(t, \mathbf{x}) = A(\mathbf{x})$. MacCann’s theorem gives us the existence of an optimal transport map T for the distance d_A , which corresponds to a constant speed displacement along a geodesic path. We can show formally, as in the isotropic case of [3] that there is a link between the speed v^* and the geodesic path from \mathbf{x} to $T(\mathbf{x})$.

Proposition 3.2. *The estimation of the optimal transportation on the Riemannian manifold (Ω, g_A) (i.e. the Wasserstein distance associate to d_A) amounts to the minimization of the geodesic displacement energy, that is to say:*

$$\int_0^1 \int_{\Omega} v^{*T} A v^* \rho^* = \inf_{(\rho, v) \in C^0} \int_0^1 \int_{\Omega} v^T A v \rho = \inf_{S \# \rho_0 = \rho_1} \int_{\Omega} d_A(\mathbf{x}, S(\mathbf{x}))^2 \rho_0(\mathbf{x}) \, d\mathbf{x},$$

Proof. We provide a formal proof, in the spirit of [3], which could be justified under some smoothness assumption on the data [25]. Let $(\rho, v) \in C^0$, and X the solution of the problem

$$\begin{cases} \partial_t y(t, \mathbf{x}) = v(t, y(t, \mathbf{x})), & \forall (t, \mathbf{x}) \in (0, 1) \times \Omega, \\ y(0, \mathbf{x}) = \mathbf{x}, & \forall \mathbf{x} \in \Omega. \end{cases}$$

We classically have the relation $\rho(t, X(t, \mathbf{x})) \det(\nabla X(t, \mathbf{x})) = \rho_0(\mathbf{x})$, i.e. $X(t, \cdot) \# \rho_0 = \rho(t, \cdot)$. Using the Jensen’s inequality, it follows that

$$\begin{aligned} & \int_0^1 \int_{\Omega} \rho(t, \mathbf{x}) v(t, \mathbf{x})^T A(\mathbf{x}) v(t, \mathbf{x}) \, d\mathbf{x} \, dt \\ &= \int_0^1 \int_{\Omega} v(t, X(t, \mathbf{x}))^T A(X(t, \mathbf{x})) v(t, X(t, \mathbf{x})) \rho(t, X(t, \mathbf{x})) \det(\nabla X(t, \mathbf{x})) \, d\mathbf{x} \, dt, \\ &= \int_0^1 \int_{\Omega} \partial_t X(t, \mathbf{x})^T A(X(t, \mathbf{x})) \partial_t X(t, \mathbf{x}) \, dt \, \rho_0(\mathbf{x}) \, d\mathbf{x}, \\ &\geq \int_{\Omega} \left(\int_0^1 (\partial_t X(t, \mathbf{x})^T A(X(t, \mathbf{x})) \partial_t X(t, \mathbf{x}))^{\frac{1}{2}} \, dt \right)^2 \rho_0(\mathbf{x}) \, d\mathbf{x}, \\ &\geq \int_{\Omega} d_A(\mathbf{x}, X(1, \mathbf{x}))^2 \rho_0(\mathbf{x}) \, d\mathbf{x}, \\ &\geq \int_{\Omega} d_A(\mathbf{x}, T(\mathbf{x}))^2 \rho_0(\mathbf{x}) \, d\mathbf{x}, \end{aligned}$$

where the last inequality stems from the fact that $X(1, \cdot)$ belongs to $\mathcal{T}(\rho_0, \rho_1)$.

Conversely, assume that T is an optimal transport map for the g_A metric. Let, for almost every $\mathbf{x} \in \Omega$, the application $X(\cdot, \mathbf{x}) \in C^1([0, 1], \Omega)$ be the unique geodesic joining \mathbf{x} to $T(\mathbf{x})$ ($X(\cdot, \mathbf{x})$ is smooth in time for a.e. $\mathbf{x} \in \Omega$). We define v as

$$\begin{cases} v(t, X(t, \mathbf{x})) = \partial_t X(t, \mathbf{x}) & \forall (t, \mathbf{x}) \in (0, 1) \times \Omega, \\ v(t, y) = 0 & \text{elsewhere,} \end{cases}$$

and $\rho(t, \cdot) = X(t, \cdot) \# \rho_0$. As a consequence, for all $\phi \in C_c^\infty((0, 1) \times \Omega)$, we obtain:

$$\begin{aligned} \int_0^1 \int_\Omega (\partial_t \phi + v \cdot \nabla \phi) \rho &= \int_0^1 \int_\Omega (\partial_t \phi(t, X(t, \mathbf{x})) + v(t, X(t, \mathbf{x})) \cdot \nabla \phi(t, X(t, \mathbf{x}))) \rho_0(\mathbf{x}) \, d\mathbf{x} \, dt \\ &= \int_\Omega \int_0^1 (\partial_t \phi(t, X(t, \mathbf{x})) + \partial_t X(t, \mathbf{x}) \cdot \nabla \phi(t, X(t, \mathbf{x}))) \rho_0(\mathbf{x}) \, dt \, d\mathbf{x} \\ &= \int_\Omega \int_0^1 \frac{d}{dt} (\phi(t, X(t, \mathbf{x}))) \, dt \, \rho_0(\mathbf{x}) \, d\mathbf{x} = 0 \end{aligned}$$

i.e. $\partial_t \rho + \operatorname{div}_{\mathbf{x}}(\rho v) = 0$ in $\mathcal{D}'((0, 1) \times \Omega)$, therefore $(\rho, v) \in C^0$. By definition we have:

$$\int_\Omega d_A(\mathbf{x}, T(\mathbf{x}))^2 \rho_0(\mathbf{x}) \, d\mathbf{x} = \int_\Omega \left(\int_0^1 (\partial_t X(t, \mathbf{x})^T A(X(t, \mathbf{x})) \partial_t X(t, \mathbf{x}))^{\frac{1}{2}} \, dt \right)^2 \rho_0(\mathbf{x}) \, d\mathbf{x}.$$

As $t \mapsto X(t, x)$ is a geodesic path and A is independent of time, $\frac{d}{dt} (\partial_t X(t, \mathbf{x})^T A(X(t, \mathbf{x})) \partial_t X(t, \mathbf{x})) = 0$, so that

$$\left(\int_0^1 (\partial_t X(t, \mathbf{x})^T A(X(t, \mathbf{x})) \partial_t X(t, \mathbf{x}))^{\frac{1}{2}} \, dt \right)^2 = \int_0^1 \partial_t X(t, \mathbf{x})^T A(X(t, \mathbf{x})) \partial_t X(t, \mathbf{x}) \, dt.$$

We finally obtain, since $(\rho, v) \in C^0$,

$$\begin{aligned} \int_\Omega d_A(\mathbf{x}, T(\mathbf{x}))^2 \rho_0(\mathbf{x}) \, d\mathbf{x} &= \int_0^1 \int_\Omega \partial_t X(t, \mathbf{x})^T A(X(t, \mathbf{x})) \partial_t X(t, \mathbf{x}) \rho_0(\mathbf{x}) \, dt \, d\mathbf{x}, \\ &= \int_0^1 \int_\Omega v(t, X(t, \mathbf{x}))^T A(X(t, \mathbf{x})) v(t, X(t, \mathbf{x})) \rho_0(\mathbf{x}) \, d\mathbf{x} \, dt, \\ &= \int_0^1 \int_\Omega v(t, \mathbf{x})^T A(\mathbf{x}) v(t, \mathbf{x}) \rho(t, \mathbf{x}) \, d\mathbf{x} \, dt, \\ &\geq \inf_{(\tilde{\rho}, \tilde{v}) \in C^0} \int_0^1 \int_\Omega \tilde{v}(t, \mathbf{x})^T A(\mathbf{x}) \tilde{v}(t, \mathbf{x}) \tilde{\rho}(t, \mathbf{x}) \, d\mathbf{x} \, dt. \quad \square \end{aligned}$$

3.2. Discretization, convexification and optimization

We now consider the discrete domain parameterized with $x \in \Omega \subset \mathbb{R}^d$ ($d = 2$ in the following) and $t \in [0, 1]$. The anisotropic optimal transportation problem reads:

$$\min_{(\rho, v) \in C^0} \frac{1}{2} \sum_{t=0}^1 \sum_{x \in \Omega} \rho(t, \mathbf{x}) v(t, \mathbf{x})^T A(\mathbf{x}) v(t, \mathbf{x}),$$

where $\rho(t, \mathbf{x}) \in \mathbb{R}$, $v(t, \mathbf{x}) = (v_1(t, \mathbf{x}), v_2(t, \mathbf{x})) \in \mathbb{R}^2$, and the 2×2 symmetric positive definite matrix $A(\mathbf{x})$ represents the anisotropy of the domain Ω . With the very same change of variable as before ($m = \rho v$), one can convexify the anisotropic transport cost as:

$$\mathcal{J}_A(\rho, m) = \sum_{t=0}^1 \sum_{\mathbf{x} \in \Omega} J_A(\rho(t, \mathbf{x}), m(t, \mathbf{x})),$$

with

$$J_A(\rho(t, \mathbf{x}), m(t, \mathbf{x})) = \begin{cases} \frac{m(t, \mathbf{x})^T A(\mathbf{x}) m(t, \mathbf{x})}{\rho(t, \mathbf{x})} & \text{if } \rho(t, \mathbf{x}) > 0 \\ 0 & \text{if } (\rho(t, \mathbf{x}), m(t, \mathbf{x})) = (0, 0) \\ +\infty & \text{otherwise} \end{cases}$$

where $m(t, \mathbf{x}) = (m_1(t, \mathbf{x}), m_2(t, \mathbf{x})) \in \mathbb{R}^2$. By taking $A(\mathbf{x}) = w(\mathbf{x})\text{Id}_2$, with $w(\mathbf{x}) \in \mathbb{R}$, we can recover the transport model on a Riemannian manifold characterized by w proposed in [30]. The anisotropic optimal transportation can therefore be estimated by minimizing the convex functional $J_A(\rho, m) + \iota_C(\rho, m)$.

In order to apply proximal splitting algorithms to this new problem, we need to compute the proximal operator of the new cost transport term. This can be done by canceling the derivative of the optimization problem defining the proximal map $\text{Prox}_{\gamma J_A}(\tilde{\rho}, \tilde{m})$. The following proposition shows that the functional J_A can still be computed in a closed form.

Proposition 3.3. *For all $(\tilde{\rho}, \tilde{m}) \in \mathbb{R} \times \mathbb{R}^d$,*

$$\text{Prox}_{\gamma J_A}(\tilde{\rho}, \tilde{m}) = \begin{cases} (\rho^*, \mu_A(\rho^*)) & \text{if } \rho^* > 0, \\ (0, 0) & \text{otherwise.} \end{cases}$$

where

$$\forall \rho \geq 0, \quad \mu_A(\rho) = (\rho \text{Id}_2 + \gamma(A + A^T))^{-1} \rho \tilde{m}$$

and ρ^* is the largest real root of the five order polynomial equation in ρ

$$P(\rho) = (\rho - \tilde{\rho}) - \gamma \tilde{m} (\rho \text{Id}_2 + \gamma(A + A^T))^{-1} A (\rho \text{Id}_2 + \gamma(A + A^T))^{-1} \tilde{m} = 0.$$

The proof is a simple extension of the one in [30]. In practice, ρ^* is obtained with the Newton’s method.

3.3. Illustration

In this section, we still consider the space-time discretization of size 32^3 and the GFB algorithm. As a first example, we present the transport of a 2D bump in an anisotropic domain, where A is defined for $\mathbf{x} = (x, y) \in [0, 1]^2$ as:

$$A_\lambda(\mathbf{x}) = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} & \text{if } x \leq 0.5, \\ \text{Id}_2 & \text{otherwise.} \end{cases} \tag{3.2}$$

The left part of the image domain ($x \leq 0.5$) then involves an anisotropic transport as soon as $\lambda \neq 1$. In fact, by setting $\lambda > 1$, this penalizes the vertical direction for the mass transfer, so that the transport will move the mass in the right part of the image for decreasing the transport cost. The results, for increasing values of $\lambda \in \{1, 3, 20\}$ are presented in Figure 2. When $\lambda = 1$, it corresponds to the classic optimal transport and we recover a translation of the bump. When the value of λ increases, we can observe that the optimal path send more and more mass to the right side of the image domain. However, note that there is presumably no splitting of mass, as we might believe looking at Figure 2, the splitting effect being induced by the fixed number of level sets which are plotted.

For the same example, we next define A as

$$A_{\lambda, \mu}(\mathbf{x}) = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} & \text{if } x \leq 0.5, \text{ and } \begin{bmatrix} 1 & 0 \\ 0 & \mu \end{bmatrix} & \text{otherwise.} \end{cases} \tag{3.3}$$

In this case, the vertical direction is favored in the right part of the image domain. The results, for $\lambda = 20$ and decreasing values of $\mu \in \{0.01, 0\}$ are presented in Figure 3. With $\mu = 0$, one can observe that a vertical teleportation of the mass arises in the right part of the domain, since the corresponding transport cost is null.

It is therefore possible to define a polarization of the space through the anisotropic matrix for modelling domain priors. This is also illustrated in Figure 4, which presents the optimal path between an horizontal and a vertical line on a square image. For these experiments, we defined different complex anisotropies, that are represented with the arrows of the first image of each line. If prior knowledge is available on the data, one can build a specific Riemannian manifold and the anisotropic model will therefore be able to simulate rigid or divergence free transports. This is nevertheless not sufficient for more general purposes and we now propose to directly include physical prior into the model and not into the domain.

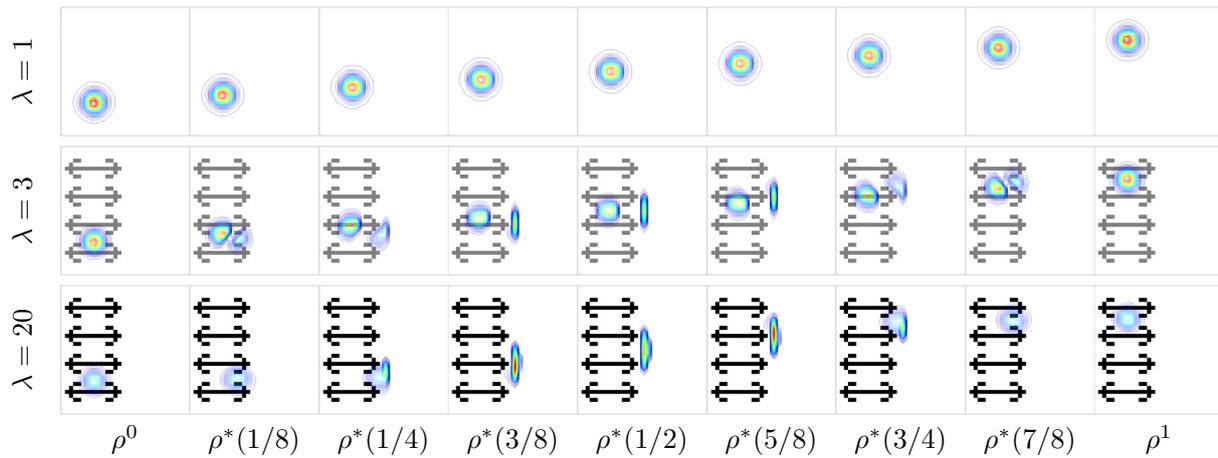


FIGURE 2. Illustration of the mass $\rho^*(t)$ estimated between ρ^0 and ρ^1 , through the computation of the generalized transport costs defined by the anisotropic matrices A_λ introduced in (3.2). The arrows represent the direction favored by the anisotropy. The color intensity denotes the strength of the anisotropy.

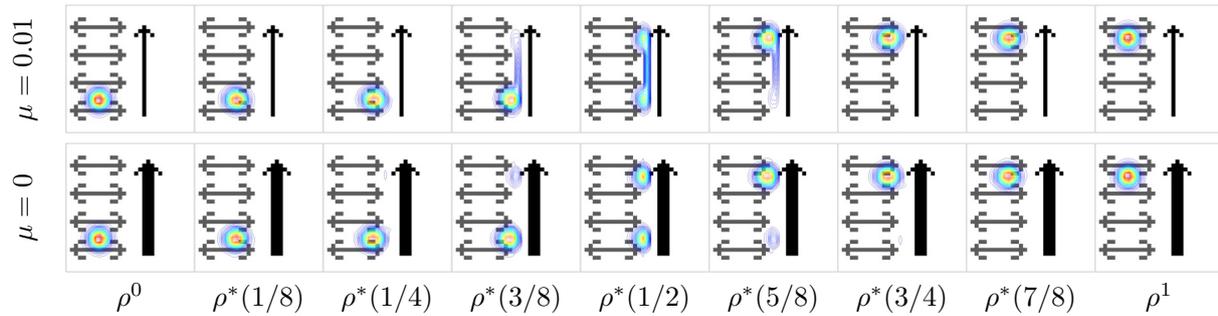


FIGURE 3. Illustration of the mass $\rho^*(t)$ estimated between ρ^0 and ρ^1 , through the computation of the generalized transport costs defined by the anisotropic matrices $A_{\lambda,\mu}$ introduced in (3.3), with $\lambda = 20$. The arrows represent the direction favored by the anisotropy. The color intensity denotes the strength of the anisotropy.

4. OPTIMAL TRANSPORTATION WITH PHYSICAL PRIORS

In this work, we are interested in incorporating physic priors into the optimal transportation model. More precisely, for image processing applications, we would like to consider incompressible or rigid transports. Such transports can indeed prevent the object contained in the data from splitting along the computed paths. They can be characterized using the velocity, that is not a variable of the problem (2.7) anymore and should be reintroduced.

4.1. Non convex coupling

In order to have a coupling between the variables (ρ, m) and v , a natural idea is to consider the term $\iota_D(\rho, m, v) = 0$, with the set $D = \{m = \rho v\}$. As D is not convex and ι_D non-smooth, we rather consider a

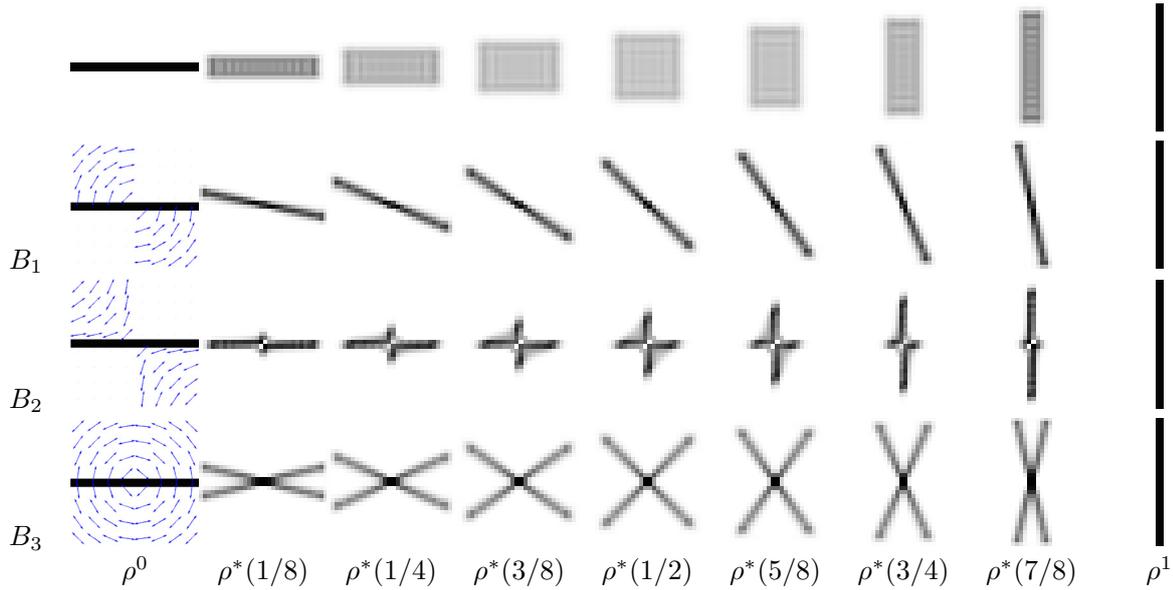


FIGURE 4. Illustration of the mass $\rho^*(t)$ estimated between ρ^0 and ρ^1 , through the computation of the transport costs defined by the three different anisotropic domains, illustrated by the blue directions in the first column. The first line is the classic isotropic transport.

differentiable coupling:

$$K(\rho, m, v) = \frac{1}{2} \int_{\Omega} \int_0^1 \|m - \rho v\|^2 \, d\mathbf{x} \, dt.$$

In this paper, we are interested in specific kind of transports regarding velocity priors that would depend on the targeted application. This will involve a new functional term $R(v)$ that can be defined as:

- Divergence-free constraint: $R(v) = \iota_{C^v}(v)$, where $C^v = \{\text{div}_{\mathbf{x}} v(t, \cdot) = 0, \langle v(t, \cdot), \mathbf{n} \rangle = 0, \forall t \in [0; 1]\}$.
- Incompressible penalization: $R(v) = \frac{1}{2} \int_0^1 \|\text{div}_{\mathbf{x}}(v(t, \cdot))\|_{L^2(\Omega)}^2 \, dt$.
- Rigid penalization: $R(v) = \frac{1}{2} \int_0^1 \|(\nabla_{\mathbf{x}} v(t, \cdot) + (\nabla_{\mathbf{x}} v(t, \cdot))^T)/2\|_{L^2(\Omega)}^2 \, dt$.
- Translation penalization: $R(v) = \frac{1}{2} \int_0^1 \|(\nabla_{\mathbf{x}} v(t, \cdot))\|_{L^2(\Omega)}^2 \, dt$.

Note that by *e.g. translation penalization*, we mean a penalization of the deviation of a velocity field from translations. These terms are convex and the three last ones are differentiable.

4.2. Existence of a minimizer

In this section we prove the existence of at least a minimizer to an optimal transportation problem with a penalization on the velocity gradient (referred as translation penalization) and the corresponding coupling term. We will assume that ρ_0 and ρ_1 belong to $L^\infty(\Omega)$, and set $M = \max(\|\rho_0\|_\infty, \|\rho_1\|_\infty)$. We set $Q = (0, 1) \times \Omega$, and denote as $BL(\Omega)$ the Beppo–Levi space on $L^2(\Omega)$ [19], that is

$$BL(\Omega) = \{u \in \mathcal{D}'(\Omega), \nabla u \in L^2(\Omega)\}.$$

Our set of constraints will be defined as:

$$C_M = \left\{ (\rho, m, v) \in L^\infty(Q) \times L^2(Q)^d \times L^2(0, 1; BL(\Omega))^d, \quad \|\rho\|_{L^\infty(Q)} \leq M, \right. \\ \left. \forall \psi \in C^\infty(Q), \int_0^1 \int_\Omega \rho \partial_t \psi + m \cdot \nabla \psi \, dx \, dt + \int_\Omega \psi(0, \mathbf{x}) \rho_0(\mathbf{x}) - \psi(1, \mathbf{x}) \rho_1(\mathbf{x}) \, dx = 0 \right\}.$$

Notice that this set is non empty since $((1 - t)\rho_0 + t\rho_1, m, 0) \in C_M$ for a constant in time m such that $\operatorname{div} m = \rho_1 - \rho_0$ and $\langle m \cdot \mathbf{n} \rangle = 0$ on $\partial\Omega \times (0, 1)$ (the existence of such m is ensured since $\int_\Omega \rho_0 \, dx = \int_\Omega \rho_1 \, dx$, see [22], p. 24).

Remark 4.1. Note that if $\rho \in C_M$, then $\int_\Omega \rho(t, \mathbf{x}) \, dx = 1$ for almost every $t \in (0, 1)$. This is readily obtained taking $\psi(t, \mathbf{x}) = t$ in the definition of C_M .

Proposition 4.2. *The problem*

$$\inf_{(\rho, m, v) \in C_M} \int_0^1 \int_\Omega J(\rho, m) + \frac{1}{2} \|m - \rho v\|^2 + \frac{1}{2} \|\nabla_{\mathbf{x}} v\|^2 \, dx \, dt$$

has at least a minimizer.

Proof. Let $(\rho_n, m_n, v_n) \in C_M$ be a minimizing sequence. As (ρ_n) is uniformly bounded by M , we have $J(m_n, \rho_n) = \frac{m_n^2}{2\rho_n} \geq \frac{1}{2M} m_n^2$. Thus we have the following *a priori* bounds: $\exists C > 0$ such that for all $n \in \mathbb{N}$,

$$\|\rho_n\|_{L^\infty(Q)} \leq M, \quad \|m_n\|_{L^2(Q)} \leq C, \quad \|\nabla_{\mathbf{x}} v_n\|_{L^2(Q)} \leq C, \quad \|m_n - \rho_n v_n\|_{L^2(Q)} \leq C. \tag{4.1}$$

The first and third estimates imply

$$\exists C > 0, \quad \|\rho_n v_n\|_{L^2(Q)} \leq C, \tag{4.2}$$

while as (ρ_n, m_n) verifies $\partial_t \rho_n + \operatorname{div}_x m_n = 0$ in $\mathcal{D}'(Q)$, we also get

$$\exists C > 0, \quad \|\partial_t \rho_n\|_{L^2(0, 1; H^{-1}(\Omega))} \leq C. \tag{4.3}$$

Therefore, up to the extraction of a subsequence, we can assume that there exists $(\rho, m, v) \in L^\infty(Q) \times L^2(Q)^d \times L^2(0, 1; BL(\Omega))^d$ such that

$$\rho_n \overset{*}{\rightharpoonup} \rho \text{ in } L^\infty(Q), \quad m_n \rightharpoonup m \text{ in } L^2(Q)^d, \quad \nabla_{\mathbf{x}} v_n \rightharpoonup \nabla_{\mathbf{x}} v \text{ in } L^2(Q)^{d \times d}. \tag{4.4}$$

First note that C_M is closed with respect to these convergences, thus the weak limit $(\rho, m, v) \in C_M$.

The integral functional \mathcal{J} is convex and lower semi-continuous for the $L^2(Q) \times L^2(Q)$ weak convergence (see [2], p. 138). The regularization term in v is the semi-norm of $L^2(0, 1; BL(\Omega))$, it is also weakly semi-continuous for the convergence of $(\nabla_{\mathbf{x}} v_n)$ in $L^2(Q)^{d \times d}$. The coupling term is also a $L^2(Q)$ norm, therefore to conclude, it remains to prove the weak convergence of $(\rho_n v_n)$ to ρv in $L^2(Q)$. We will use a form of compensated compactness lemma, using the bound on the temporal derivative for ρ_n and on spatial derivatives for v_n , as in [20]. In this spirit, the following is a mere corollary of a lemma of [28]:

Lemma 4.3. *Consider (a_n) weakly convergent to a in $L^2(0, 1; H^1(\Omega))$, and (b_n) weakly convergent to b in $L^2(Q)$. If $(\partial_t b_n)$ is bounded in $L^2(0, 1; H^{-1}(\Omega))$ then up to a subsequence, $(a_n b_n)$ converges to ab in $\mathcal{M}(Q)$ (that is for test functions in $\mathcal{C}_c^0(Q)$).*

One issue here is that we have no control on v_n in the regularization term, but only on $\nabla_{\mathbf{x}} v_n$, hence the space $BL(\Omega)$ instead of the required $H^1(\Omega)$. Hopefully we will get this control from the coupling term, which is not trivial since ρ_n is only nonnegative.

From [19], page 312, since Q is star-shaped, the weak convergence of (v_n) to v in $L^2(0, 1; BL(\Omega))$ implies the weak convergence of $(v_n + c_n)$ to v in $L^2(Q)$ for some constant in space sequence $(c_n(t))$.

As ρ_n is bounded in $L^\infty(Q)$, the sequence $\rho_n(v_n + c_n)$ is in fact bounded in $L^2(Q)$, and thanks to (4.2) we get that $\rho_n c_n$ is bounded in $L^2(Q)$. By definition,

$$\|\rho_n c_n\|_{L^2(Q)} = \sup_{\psi \in L^2(Q), \|\psi\|_{L^2(Q)}=1} \int_0^1 \int_\Omega \rho_n c_n \psi \, d\mathbf{x} \, dt,$$

which, restricted to constant in space functions $\psi \in L^2(0, 1)$ in the supremum, gives

$$\|\rho_n c_n\|_{L^2(Q)} \geq \sup_{\psi \in L^2(0,1), \|\psi\|_{L^2(0,1)}^2 = \frac{1}{|\Omega|}} \int_0^1 \int_\Omega \rho_n c_n \psi \, d\mathbf{x} \, dt.$$

Moreover, as c_n is constant in space too, and since $\int_\Omega \rho_n \, d\mathbf{x} = 1$ (see Rem. 4.1), we get

$$\|\rho_n c_n\|_{L^2(Q)} \geq \sup_{\psi \in L^2(0,1), \|\psi\|_{L^2(0,1)}^2 = \frac{1}{|\Omega|}} \int_0^1 c_n \psi \, dt = \frac{1}{\sqrt{|\Omega|}} \|c_n\|_{L^2(0,1)}.$$

This means that (c_n) is bounded in $L^2(0, 1)$ and also in $L^2(Q)$ as a (constant in \mathbf{x}) sequence of this space. It now implies that v_n is bounded in $L^2(Q)$, and therefore converges up to a subsequence to some function w . Necessarily from (4.4) we have $w = v$ since ∇ is weakly continuous. Thus using estimates (4.1) and (4.3), we can apply the above compactness lemma to v_n and ρ_n , to obtain

$$\rho_n v_n \rightarrow \rho v \text{ in } \mathcal{M}(Q).$$

However $\rho_n v_n$ is bounded in $L^2(Q)$ weak, and therefore a subsequence weakly converges in this space. Extracting another subsequence if necessary we get

$$\rho_n v_n \rightarrow \rho v \text{ in } L^2(Q).$$

□

Remark 4.4. Concerning other penalizations, while we will see that numerical algorithms behave as expected, the situation from the theoretical point of view seems more tricky. The case of rigid penalization, while conceptually very close to the above one, thanks to Korn inequality, has to be studied carefully since the simplification obtained with a constant c_n does not generalize directly to a rigid one. An easy case would be of course to add a $L^2(Q)$ norm on v with a small positive coefficient in the functional. The case of incompressible penalization would moreover need an extension of the compactness lemma to that case.

4.3. Discretization

We will consider the same discretizations as before and now detail the case $d = 2$. We assume that the centered grid is of size $P \times N \times N$: P points for the temporal dimension and N points for each spatial one. The velocity $w = (w^1, w^2)$ on the centered grid is discretized as

$$w \in \mathcal{E}_c^v = (w_{k,i,j}^1, w_{k,i,j}^2)_{\substack{0 \leq k \leq P \\ 0 \leq i, j \leq N}},$$

where $\mathcal{E}_c^v = (\mathbb{R}^d)^{\mathcal{G}}$. Hence, a velocity $v = (v^1, v^2)$ is defined on the staggered grid as

$$v \in \mathcal{E}_s^v = \left((v_{k,i,j}^1)_{\substack{0 \leq j \leq N \\ -1 \leq i \leq N}}, (v_{k,i,j}^2)_{\substack{-1 \leq j \leq N \\ 0 \leq i \leq N}} \right)_{0 \leq k \leq P},$$

where $\mathcal{E}_s^v = \mathbb{R}^{\mathcal{G}^x}$ is the finite dimensional space of staggered variables.

We also introduce \mathcal{I}^v as a midpoint interpolation operator from \mathcal{E}_s^v to \mathcal{E}_c^v , defined in a similar way to \mathcal{I} . Let us finally detail the previously introduced spatial divergence and gradient operators in the case $d = 2$ where $v = (v^1, v^2)$. The spatial divergence operator $\operatorname{div}_{\mathbf{x}} : \mathcal{E}_s^v \rightarrow \mathbb{R}^{\mathcal{G}}$ is defined, for $v \in \mathcal{E}_s^v$ as

$$\forall 0 \leq k \leq P, \forall 0 \leq i \leq N, \forall 0 \leq j \leq N, \quad \operatorname{div}_{\mathbf{x}}(v)_{k,i,j} = (v_{k,i,j}^1 - v_{k,i-1,j}^1) + (v_{k,i,j}^2 - v_{k,i,j-1}^2).$$

Finally, the spatial gradient operator is defined, for $v = (v^1, v^2) \in \mathcal{E}_s^v$ as the matrix $\nabla_{\mathbf{x}} v = [\nabla_{\mathbf{x}_i}(v^{\tilde{j}})]_{i=1,2}^{\tilde{j}=1,2}$.

In this matrix, $\nabla_{\mathbf{x}_i}(v^{\tilde{j}})$ is the partial derivative of $v^{\tilde{j}}$ on the dimension $\mathbf{x}_i \in (x, y)$. We can observe that $\nabla_{\mathbf{x}_1}(v^1), \nabla_{\mathbf{x}_2}(v^2) \in \mathbb{R}^{\mathcal{G}}$, whereas $\nabla_{\mathbf{x}_2}(v^1)$ and $\nabla_{\mathbf{x}_1}(v^2)$ are defined on double staggered grids of respective dimensions $P \times (N + 1) \times (N - 1)$ and $P \times (N - 1) \times (N + 1)$.

4.4. Optimization

The generalized optimal transport model we are interested in is of the following problem:

$$(\rho^*, m^*, v^*) = \operatorname{argmin}_{\rho, m, v} F(\rho, m, v) \tag{4.5}$$

where

$$F(\rho, m, v) = \mathcal{J}(\mathcal{I}(\rho, m)) + \iota_C(\rho, m) + \lambda K(\mathcal{I}(\rho, m), \mathcal{I}^v(v)) + \alpha R(v),$$

and $\lambda, \alpha \geq 0$ respectively weight the coupling between variables and the velocity regularization term.

The problem (4.5) is not convex in (ρ, m, v) but it is convex separately in (ρ, m) and in v . Notice that the coupling is differentiable and the non-smooth terms are separable. Through these assumptions and following [33], we can perform a block coordinate descent and minimize alternatively each convex problem to obtain a critical point of the joint problem (4.5). The optimization algorithm reads:

$$\begin{cases} (\rho^{k+1}, m^{k+1}) = \operatorname{argmin}_{(\rho, m)} F(\rho, m, v^k) \\ v^{k+1} = \operatorname{argmin}_v F(\rho^{k+1}, m^{k+1}, v). \end{cases} \tag{4.6}$$

In order to deal with the different numerical grids, we define again a variable $x = (\bar{\rho}, \bar{m})$ on the regular grid and a variable y on the staggered one, for the couple (ρ, m) . We do the same for the velocity and denote as w the variable on the centered grid and v the velocity on the staggered one.

We finally want to solve the problem

$$(x^*, y^*, v^*, w^*) = \operatorname{argmin}_{x, y, v, w} \tilde{F}(x, y, v, w) \tag{4.7}$$

where

$$\tilde{F}(x, y, v, w) = \mathcal{J}(x) + \iota_C(y) + \iota_S(x, y) + \lambda K(x, w) + R(v) + \iota_{S^v}(v, w),$$

and the convex set S^v is defined for the velocity variables as $S^v(v, w) = \{\mathcal{I}^v(v) = w\}$. A critical point of (4.7) can therefore be estimated with the following algorithm

$$\begin{cases} (x^{k+1}, y^{k+1}) = \operatorname{argmin}_{(x, y)} \tilde{F}(x, y, w^k), \\ (v^{k+1}, w^{k+1}) = \operatorname{argmin}_{(v, w)} \tilde{F}(x^{k+1}, v, w), \end{cases} \tag{4.8}$$

that is composed of alternated optimization problems. The resolution of these two problems is now detailed.

4.4.1. Optimization of (ρ, m)

The first step of (4.8) reads:

$$(x^*, y^*) = \underset{x, y}{\operatorname{argmin}} \mathcal{J}(x) + \iota_C(y) + \iota_S(x, y) + \lambda K(x, w^k). \tag{4.9}$$

We can again use the GFB algorithm 1 to solve this problem, by taking $F(x, y) = K(x, w^k)$, $G_1(x, y) = \mathcal{J}(x) + \iota_C(y)$ and $G_2(x, y) = \iota_S(x, y)$. The Lipschitz constant of ∇F is given by $L = 2\lambda \max(1, \|w^k\|_\infty^2)$. As a consequence, when the coupling between variables is enforced with a large value of λ , it decreases the value of the time-step of the optimization algorithm τ , that is inversely proportional to L .

4.4.2. Optimization of v

Let us now detail how dealing with the different regularization models of the velocity we have in mind.

Divergence-free constraint

When the velocity v is constrained to be incompressible, the second step of (4.8) reads:

$$(v^*, w^*) = \underset{v, w}{\operatorname{argmin}} \lambda K(x^{k+1}, w) + \iota_{C^v}(v) + \iota_{S^v}(w, v). \tag{4.10}$$

The GFB Algorithm 1 can be applied to solve this problem by taking $F(v, w) = \lambda K(x^{k+1}, w)$, $G_1(v, w) = \iota_{C^v}(v)$ and $G_2(v, w) = \iota_{S^v}(w, v)$. The Lipschitz constant of ∇F is then given by $L = \lambda \|\bar{\rho}^{k+1}\|_\infty^2$.

The computation of $\operatorname{Prox}_{\iota_{C^v}}(\tilde{v})$ corresponds to the projection of $\tilde{v}(t, \cdot, \cdot)$, on the $2D$ divergence free velocity fields $\forall t \in [0; 1]$. At each time t , we first extract the boundary values of the velocity on the staggered grid using the linear operator b^v , defined for $d = 2$, $\Omega = [0; 1]^2$ and $\tilde{v} = (\tilde{v}^1, \tilde{v}^2)$ as:

$$b^v(\tilde{v}(t, \cdot, \cdot)) = (\tilde{v}^1(t, 0, \cdot), \tilde{v}^1(t, 1, \cdot), \tilde{v}^2(t, \cdot, 0), \tilde{v}^2(t, \cdot, 1)).$$

Denoting the homogeneous Neumann boundary conditions as $b_0^v = (0, 0, 0, 0)$, the affine set C^v that can be rewritten as

$$C^v = \{v, B_t v(t, \cdot, \cdot) = z, \forall t \in [0; 1]\}$$

where

$$B_t v(t, \cdot, \cdot) = (\operatorname{div}_x(v(t, \cdot, \cdot)), b^v(v(t, \cdot, \cdot))), \text{ and } z = (0, b_0^v).$$

The projection v of \tilde{v} can again be computed by solving a linear system $\forall t \in [0; 1]$:

$$v(t, \cdot, \cdot) = \operatorname{Proj}_{C^v}(\tilde{v}(t, \cdot, \cdot)) = (\operatorname{Id} - B_t^* \Delta_t^{-1} B_t) \tilde{v}(t, \cdot, \cdot) + B_t^* \Delta_t^{-1} z$$

where applying $\Delta_t^{-1} = (B_t B_t^*)^{-1}$. Following the expression of $\operatorname{Prox}_{\tau \iota_S}$ defined in (2.11), we obtain in a similar way $\operatorname{Prox}_{\tau \iota_{S^v}}$ by using \mathcal{I}^v instead of \mathcal{I} .

Penalization

If we want to penalize a certain norm of the velocity with a term $\alpha R(v)$, where R is proper, convex and differentiable with a Lipschitz gradient, we obtain the following problem:

$$(v^*, w^*) = \underset{v, w}{\operatorname{argmin}} \lambda K(x^{k+1}, w) + \alpha R(v) + \iota_{S^v}(w, v), \tag{4.11}$$

that can be solved with the FB Algorithm 2, by taking $G(v, w) = \iota_{S^v}(w, v)$ and $F(v, w) = \lambda K(x^{k+1}, w) + \alpha R(v)$. The Lipschitz constant of ∇F is given by $L = \lambda \|\bar{\rho}^{k+1}\|_\infty^2 + \alpha L_R$, where L_R is the lipschitz constant of ∇R . For an incompressible penalization $R(v) = \frac{1}{2} \sum_{t=0}^1 \|\operatorname{div}_x(v(t, \cdot, \cdot))\|_{L^2(\Omega)}^2$, as for a rigid penalization $R(v) = \frac{1}{2} \sum_{t=0}^1 \|(\nabla_x v(t, \cdot, \cdot) + (\nabla_x v(t, \cdot, \cdot))^T)/2\|_{L^2(\Omega)}^2$ or a translation penalization $R(v) = \frac{1}{2} \sum_{t=0}^1 \|\nabla_x v(t, \cdot, \cdot)\|_{L^2(\Omega)}^2$, we have $L_R = 4\sqrt{2}$ with our discrete gradient and divergence operators computed on regular and staggered grids with coefficients $[1, -1]$.

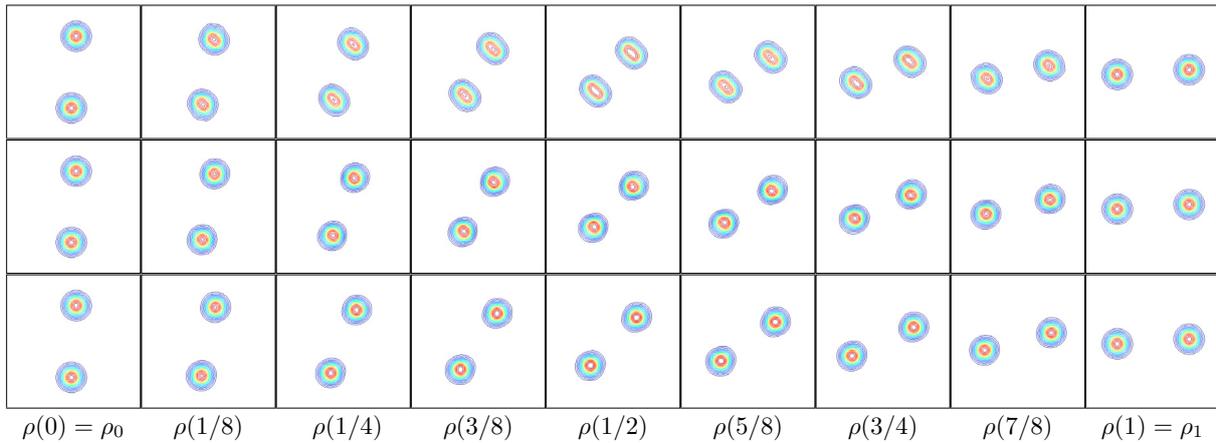


FIGURE 5. Two gaussians experiments with penalization and a null initialization. Plot of the isolevels of the density $\rho(t)$ along the optimal path computed with the different penalization models. The top line is realized with incompressibility penalization, the second with a translation penalization (through the minimization of the L^2 norm of each component of the velocity field) and the last one with a rigid penalization.

5. EXPERIMENTS

We now illustrate the influence of the regularization and penalization of the velocity field on different examples. In all our experiments, we take $\lambda = 10^3$ for the coupling term and $\alpha = 2 \times 10^3$ for the regularization of the velocities. The time-step of the GFB algorithm is therefore taken as $\tau = 10^{-4}$. In practice, depending of the dimension of the problem, we realize until 1000 outer iterations of the alternated algorithm (4.8) and 10 inner iterations of both problems (4.9) and (4.11), when a penalization of the velocity is involved. Notice that only a single step is required for the model (4.10) through a projection of v on the divergence-free set C^v . Such parameters are sufficient to have a convergence of the transport costs and coherent computed paths.

5.1. Synthetic tests

First of all, we compare in Figure 5 the results obtained with the incompressible, translation and rigid penalizations on the two Gaussians example. Such physical priors then prevent the mass from splitting along the computed path. The Gaussians are deformed with the divergence-free prior, but it can be seen that the length of the level lines of the densities are preserved along the path. It is also important to underline that both translation and rigid penalizations keep the exact shapes of the two Gaussians along time. Furthermore, one can see in Figure 6 representing both computed paths, that the rigid penalization really performs a rotation and not a translation, so that the optimal path is no more composed of straight lines.

In the example of Figure 7 that presents a rotating bar, the rigid penalization (last line) recovers a quasi-rotation, which better preserves the prior physics with respect to pure optimal transport (first line). As expected, it can also be observed in Figure 8 that the length of the level lines of the estimated density are preserved with the incompressible and rigid penalization approaches. We refer the reader to [9] for more synthetic examples involving such penalizations.

On these synthetic examples, we considered a discrete spatio-temporal grid of 31^3 points. For these dimensions, only 100 outer iterations of the block coordinate algorithm (4.8) are needed.

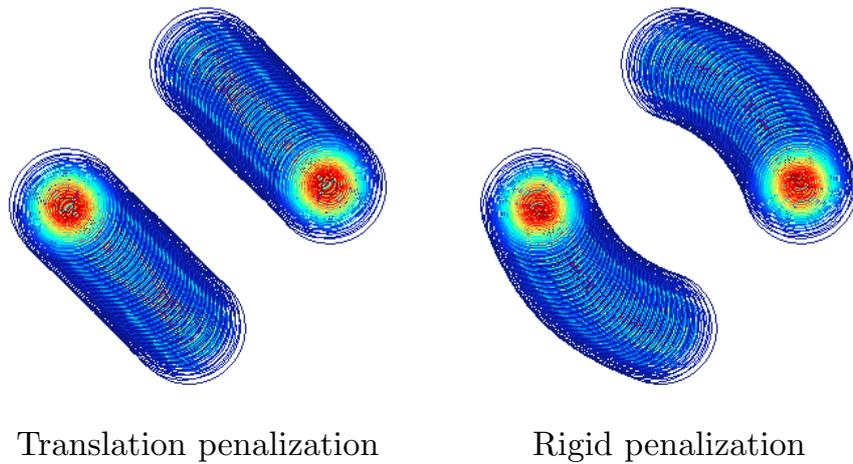


FIGURE 6. Two gaussians experiment. Plot of the whole trajectory computed with the translation (*on the left*) and the rigid (*on the right*) penalization models. The rigidity here involves a real rotation.

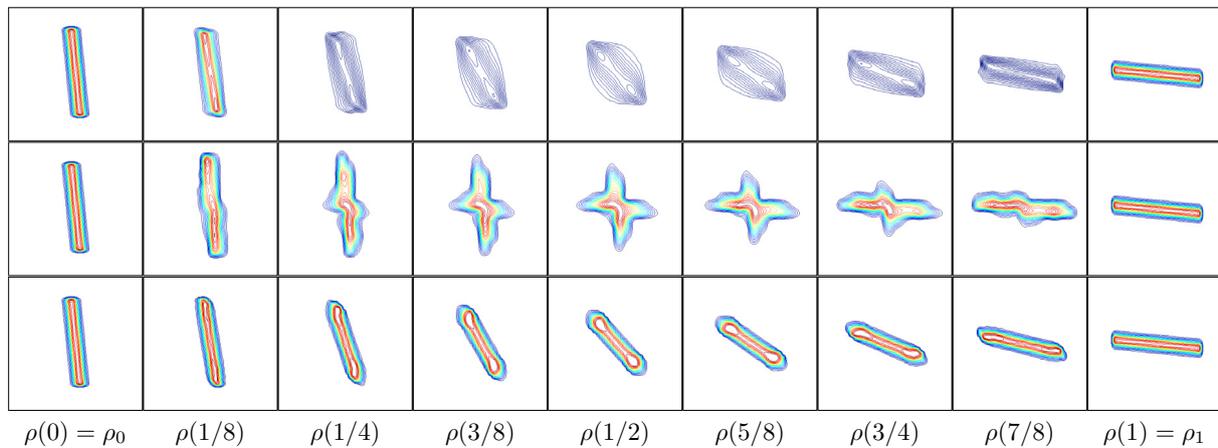


FIGURE 7. Bar experiment. Plot of the isolevels of the density $\rho(t)$ along the optimal path computed with different approaches. The first line is the optimal transport of [3], the second is the proposed approach with an incompressible penalization and the last one with a rigid penalization that conserves the nature of the object to transport.

5.2. Image interpolation in oceanography

In order to study the state of the oceans, snapshot images are produced by operational numerical codes such as the Ocean Circulation Model NEMO³. The synthesis process is nevertheless time-consuming and the oceanographers would like to create a few images and then realize a temporal interpolation between these images. This would be of particular interest to visualize and diffuse movies of dynamic structures of the ocean. The main issue comes from the coast that appears in a lot of interesting places, as illustrated in Figure 9.

³<http://www.nemo-ocean.eu/>

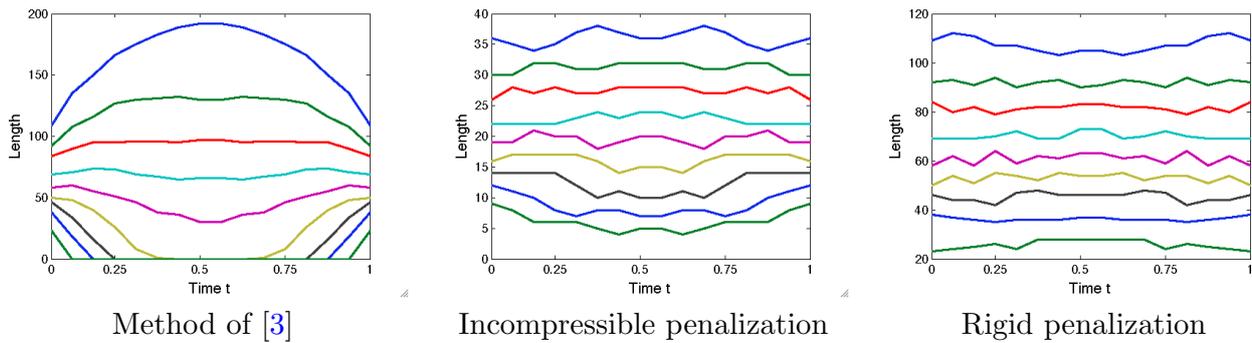


FIGURE 8. Bar experiment. Evolution of the length of the upper level lines of the estimated density along time t : $|\rho(t, \mathbf{x}) > i/10|$, for $i = 1 \dots 9$. The left plot is the classic optimal transport of [3], the middle one is the proposed approach with an incompressible penalization and the right one corresponds to the proposed approach with a rigid penalization. Penalizing the norm of the velocity makes the level lines preserved along the computed path.

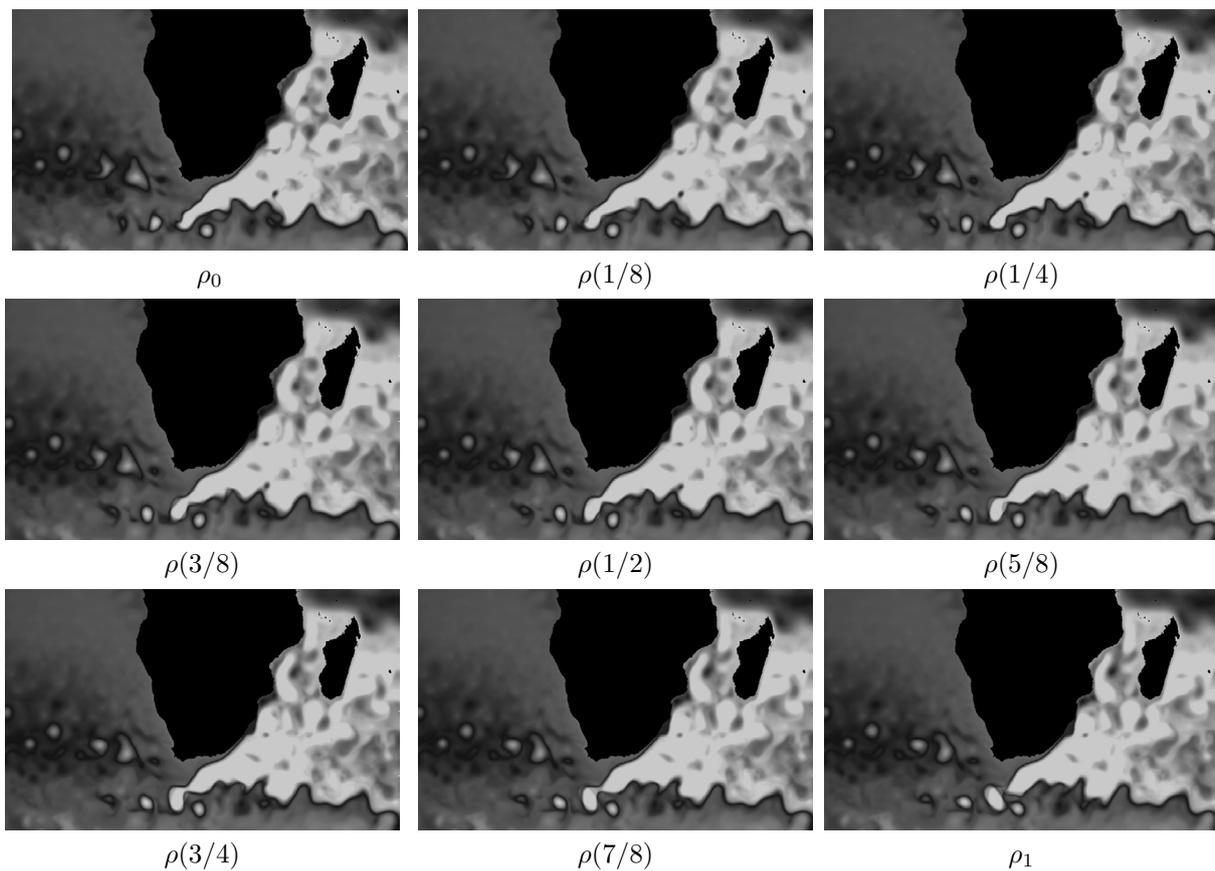


FIGURE 9. Interpolation of Sea Surface Height images in oceanography: creation of vortices in Cap Point.

It makes useless classical image registration techniques such as optical flow or diffeomorphism estimations that can not deal with such complex domain.

With our anisotropic formulation, we can represent the image domain as a Riemannian manifold, by taking $A(t, \mathbf{x}) = w(\mathbf{x}) \text{Id}_2$. The variable $w(\mathbf{x})$ then describes the manifold, it is set to 1 in the ocean and to a huge value in the land in order to restrict the transport into the ocean.

The successive optimal paths between 10 pairs of Sea Surface Height images of size 843×516 produced by the NEMO model have thus been computed with a temporal discretization of $P = 9$ steps⁴. The brightest colors correspond to the highest sea height. The image sequence illustrates the Agulhas Current and the creation of vortexes in Cap Point. As shown in Figure 9, that presents the computed path between two consecutive images, by adding the proposed divergence-free penalization, we can induce some rotational prior within the transport estimation and perfectly recover the creation of vortexes. In this experiment, we considered 1000 outer iterations of the alternated process (4.8).

6. CONCLUSION

In this paper, we study generalized optimal transportation models which attach a multiphysics model to the images to be registered. This is of particular interest for image interpolation purposes, where the results obtained using a simple minimization of a kinetic energy under some constraints do not preserve image characteristics along the optimal interpolation path, which is not physical. We then provide adequate proximal splitting methods to solve these new optimization problems. These generalizations are not limited by the expressions we consider here. Indeed, others physical terms can also be considered taking into account more complex physics.

Promising results have been obtained on high-resolution oceanographic images. Future works, in collaboration with oceanographers, will therefore be dedicated to the modelling of more accurate physical priors.

From the optimization point of view, we are also focused on the minimization of non-convex and non-smooth functionals. Contrary to the block coordinate descent method we consider (4.6), defining an algorithm without inner loops (on (ρ, m) and v) would be of main interest to obtain a faster computation of minima of our non-convex problems. Recent advances of proximal splitting methods have been made for related problems [1], but they are still limited to more simple functionals and can not deal with our generalized optimal transport formulation.

Finally, for the anisotropic optimal transport, we plan to study the existence of minimizers when the anisotropy varies with respect to time. This is a challenging problem as, in this case, there are no correspondences with any Monge problem.

Acknowledgements. The authors would like to thank Jean-Marc Molines (MEOM Team, Laboratoire de glaciologie et géophysique de l'environnement, Grenoble, France) for providing us the image sequences of sea surface height that have been created in the context of the DRAKKAR project (<http://www.drakkar-ocean.eu>). This work is supported by the French Agence Nationale de la Recherche, ANR Project <http://tommi.imag.fr> TOMMI (ANR 2011 BS01 014 01).

REFERENCES

- [1] H. Attouch, J. Bolte and B. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward backward splitting, and regularized gauss-seidel methods. *Math. Program.* **137** (2013) 91–129.
- [2] H.H. Bauschke and P.L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces. Springer (2011).
- [3] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.* **84** (2000) 375–393.
- [4] J.-D. Benamou, B.D. Froese and A.M. Oberman, Two numerical methods for the elliptic Monge-Ampère equation. *ESAIM: M2AN* **44** (2010) 737–758.

⁴The full data and results are available online: (http://www.math.u-bordeaux1.fr/~npapadak/OT/original_data.gif) and (<http://www.math.u-bordeaux1.fr/~npapadak/OT/ot.gif>)

- [5] J.-D. Benamou, B.D. Froese and A.M. Oberman, Numerical solution of the optimal transportation problem via viscosity solutions for the Monge-Ampère equation. *CoRR*, abs/1208.4873 (2012).
- [6] J.-D. Benamou, B.D. Froese and A.M. Oberman, Numerical solution of the optimal transportation problem using the Monge-Ampère equation. *J. Comput. Phys.* **260** (2014) 107–126.
- [7] N. Bonnotte, From Knothe’s rearrangement to Brenier’s optimal transport map. *SIAM J. Math. Anal.* **45** (2013) 64–87.
- [8] R.I. Bot, E.R. Csetnek and A. Heinrich, On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems. *CoRR*, abs/1303.2875 (2013).
- [9] A. Bouharguane, E. Maitre, E. Oudet and N. Papadakis, Multiphysics optimal transportation and image analysis. Technical Report hal-00740671 (2012).
- [10] A. Bouharguane, An. Iollo and L. Weynans, Numerical solution of the Monge–Kantorovich problem by Picard iterations. Technical Report RR-8477, INRIA (2014).
- [11] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.* **44** (1991) 375–417.
- [12] C. Brune, *4D imaging in tomography and optical nanoscopy*. Ph.D. thesis, University of Münster, Germany (2010).
- [13] G. Carlier, A. Galichon and F. Santambrogio, From knothe’s transport to brenier’s map and a continuation method for optimal transport. *SIAM J. Math. Anal.* **41** (2010) 2554–2576.
- [14] A. Chambolle, Rates of convergence for accelerated primal-dual algorithms. Presented in SIAM Conference on Imaging Sciences (2014).
- [15] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision* **40** (2011) 120–145.
- [16] P. Clarysse, B. Delhay, M. Picq and J. Pousin, Optimal extended optical flow subject to a statistical constraint. *J. Comput. Appl. Math.* **234** (2010) 1291–1302.
- [17] L. Condat, A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.* **158** (2013) 460–479.
- [18] E.J. Dean and R. Glowinski, An augmented Lagrangian approach to the numerical solution of the dirichlet problem for the elliptic Monge-Ampère equation in two dimensions. *Electron. Trans. Numer. Anal.* **22** (2006) 71–96.
- [19] J. Deny and J.-L. Lions, Les espaces du type de beppo levi. *Ann. Inst. Fourier* **5** (1954) 305–370.
- [20] R.J. DiPerna and P.-L. Lions, Ordinary differential equations, transport theory and sobolev spaces. *Inv. Math.* **98** (1989) 511–547.
- [21] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. Elsevier Science (1983).
- [22] V. Girault and P.A. Raviart, *Finite element methods for Navier-Stokes equations: theory and algorithms*. Springer series in computational mathematics. Springer–Verlag (1986).
- [23] S. Haker, L. Zhu, A. Tannenbaum and S. Angenent, Optimal mass transport for registration and warping. *Int. J. Comput. Vision* **60** (2004) 225–240.
- [24] F.H. Harlow and J.E. Welch, Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. *Phys. Fluids* **8** (1965) 2182–2189.
- [25] R. Hug, E. Maitre and N. Papadakis, On the convergence of augmented lagrangian method for optimal transport between nonnegative densities. Technical report, hal-01128793 (2015).
- [26] A. Iollo and D. Lombardi, A lagrangian scheme for the solution of the optimal mass transfer problem. *J. Comput. Phys.* **230** (2011) 3430–3442.
- [27] R.J. McCann, Polar factorization of maps on riemannian manifolds. *Geom. Funct. Anal. GAFA* **11** (2001) 589–608.
- [28] A. Moussa, Some variants of the classical aubin-lions lemma. preprint [arXiv:1401.7231](https://arxiv.org/abs/1401.7231) (2014).
- [29] V.I. Oliker and L.D. Prussner, On the numerical solution of the equation $\frac{\partial^2 z}{\partial x^2} \frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 = f$ and its discretizations, i. *Numer. Math.* **54** (1989) 271–293.
- [30] N. Papadakis, G. Peyré and É. Oudet, Optimal transport with proximal splitting. *SIAM J. Imaging Sci.* **7** (2014) 212–238.
- [31] H. Raguét, J. Fadili and G. Peyré, A generalized forward-backward splitting. *SIAM J. Imaging Sci.* **6** (2013) 1199–1226.
- [32] G. Strang, *Linear Algebra and Its Applications*. Brooks Cole (1988).
- [33] P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109** (2001) 475–494.
- [34] C. Villani, *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer (2008).