HAL open science

Optimal Transport with Proximal Splitting

Nicolas Papadakis, Gabriel Peyré, Edouard Oudet

To cite this version:

Nicolas Papadakis, Gabriel Peyré, Edouard Oudet. Optimal Transport with Proximal Splitting. 2013. hal-00816211v1

HAL Id: hal-00816211

https://hal.science/hal-00816211v1

Submitted on 21 Apr 2013 (v1), last revised 22 Oct 2013 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

OPTIMAL TRANSPORT WITH PROXIMAL SPLITTING

NICOLAS PAPADAKIS*, GABRIEL PEYRÉ†, AND EDOUARD OUDET‡

Abstract. This article reviews the use of first order convex optimization schemes to solve the discretized dynamic optimal transport problem, initially proposed by Benamou and Brenier. We develop a staggered grid discretization that is well adapted to the computation of the L^2 optimal transport geodesic between distributions defined on a uniform spatial grid. We show how proximal splitting schemes can be used to solve the resulting large scale convex optimization problem. A specific instantiation of this method on a centered grid corresponds to the initial algorithm developed by Benamou and Brenier. We also show how more general cost functions can be taken into account and how to extend the method to perform optimal transport on a Riemannian manifold.

1. Introduction. Optimal transport is a well developed mathematical theory that defines a family of metrics between probability distributions [54]. These metrics measure the amplitude of an optimal displacement according to a so-called ground cost defined on the space supporting the distributions. The resulting distance is sometimes referred to as the Wasserstein distance in the case of L^p ground costs. The geometric nature of optimal transportation, as well as the ability to compute optimal displacements between densities, make this theory progressively mainstream in several applicative fields such as economic modeling and image processing. However, the numerical resolution of the optimal transportation problem raises several challenges. This article is focused on the computation of geodesics for the optimal transport metric associated to the L^2 cost. It reviews and extends the approach pioneered by Benamou and Brenier [7] from the perspective of proximal operator splitting in convex optimization. This shows the simplicity and efficiency of this method, which can easily be extended beyond the setting of optimal transport by considering various convex cost functions.

1.1. Previous Works.

Discrete optimal transport. The easiest way to discretize and compute numerically optimal transports is to consider finite sums of weighted Diracs. In this specific case, the optimal transport is a multi-valued map between the Diracs locations. Specific linear solvers can be used in this context, and in particular network and transportation simplexes [22] can scale up to a few thousands of Dirac masses. Dedicated combinatorial optimization methods have been proposed, such as the auction algorithm [10], which can handle integer costs between the Diracs. In the even more restricted case of two distributions with the same number of Diracs with equal weights, the transportation is a bijection between the points, and thus corresponds to the optimal assignment problem [16]. Combinatorial optimization methods such as the Hungarian algorithm [32] have roughly cubic complexity in the number of Diracs. Faster schemes exist for specific cost functions, such as for instance convex cost of the distance on the line (where it boils down to a sorting of the positions) and the circle [25], concave costs on the line [26], the ℓ^1 distance [39]. The computation can be accelerated using multi-scale clustering [42]. Note also that various approximations of the transportation distance have been proposed, see for instance [50].

Despite being numerically intensive for finely discretized distributions, this discrete transport framework have found many applications, such as for instance color transfer between images [46], shape retrieval [48], surface reconstruction [23] and interpolation for computer graphics [11]

Optimal transport via PDE's. The optimal transport for the L^2 ground cost has a special structure. It can be shown to be uniquely defined and to be the gradient of a convex function [13]. This implies that it is also the solution of the fully non-linear Monge-Ampère partial differential equation.

Several methods have been proposed to discretize and solve this PDE, such as for instance the method of [45] which converges to the Aleksandrov solution, and the one of [44] which converges to the viscosity solution of the equation. Alternative methods such as [24] and [29] are efficient for

*IMB, Université Bordeaux 1, 351, cours de la Libération, F-33405 TALENCE, FRANCE.

†CEREMADE, Université Paris-Dauphine, Place du Marechal De Lattre De Tassigny, 75775 PARIS CEDEX 16, FRANCE.

‡LJK, Université de Grenoble, 51 rue des Mathématiques, Campus de Saint Martin d'Hères, BP 53, 38041 GRENOBLE CEDEX 09

regular densities. A major difficulty in these approaches is to deal with compactly supported densities, which requires a careful handing of the boundary conditions. [31] proposes to enforce these conditions by iteratively solving a Monge-Ampere equation with Neumann boundary conditions. [9] introduces a method requiring the solution of a well-posed Hamilton-Jacobi equation.

Another line of methods iteratively constructs mass preserving mappings converging to the optimal transport [3]. This explicitly constructs the so-called polar factorization of the initial map, see also [5] for a different approach. This method is enhanced in [33] to avoid drifting from the preservation constraint during the iterations.

These PDE's based approaches to the resolution of the optimal transport have found several applications, such as image registration [34], density regularization [15], optical flow [19] and grid generation [52].

Dynamical optimal transport. Instead of computing directly the transport, it is possible to consider the geodesic path between the two densities according to the Wasserstein metric (the so-called displacement interpolation). For the L^2 ground cost, this geodesic is obtained by linear interpolation between the identity and the transport. The geodesic can thus be computed by first obtaining the transport and then evolving the densities. If one considers discrete sums of Diracs, this corresponds to solving a convex linear program, and can also be understood as a Lagrangian approximation of the transport between (possibly continuous) densities that have been discretized. This approach is refined in [36], which considers discretization with mixture of Gaussians.

It is also possible to consider an Eulerian formulation of the geodesic problem, for which densities along the path are discretized on a fixed spatial grid. Conservation of mass is achieved by introducing an incompressible velocity field transporting the densities. The breakthrough paper [7] shows that it is possible to perform a change of variable to obtain a convex problem. They propose to solve numerically the discretized problem with a first order iterative method (more precisely, they use the ADMM algorithm on a dual formulation, see bellow).

Geodesics between pairs of distributions can be extended to barycenters between an arbitrary finite collection of distributions. Existence and uniqueness of this barycenter is studied in [1]. Computing the barycenter between discrete distributions requires the resolution of a convex linear program that corresponds to a multi-marginal optimal transportation. However, in sharp contrast with the case of two distributions, the special case of un-weighted sums of Diracs is not anymore equivalent to an assignment problem, which is known to be NP-hard [16]. Computing numerically this barycenter for large scale problems thus requires the use of a non-convex formulation to solve for a Lagrangian discretization, which finds applications in image processing [47].

Generalized transport problems. The formulation of the geodesic computation as a convex optimization problem initiated by [7] enables the definition of various metrics obtained by changing the objective function. A penalization of the matching constraint [6] allows one to compute an unbalanced transport where densities are not normalized to have the same mass. An interpolation between the L^2-Wasserstein and L^2 distances is proposed in [8]. Lastly, an interpolation between L^2-Wasserstein and H^{-1} distances is described in [27]. This extension relies in a crucial manner on the convexity of the extended objective function, which enables a theoretical analysis to characterize minimizing geodesics [17]. Convexity also allows one to use the numerical scheme we propose with only slight modifications with respect to the L^2-Wasserstein case.

Optimal transport on Riemannian manifolds. Many properties of the L^2-Wasserstein distance extend to the setting where the ground cost is the square of the geodesic distance on a Riemannian manifold. This includes in particular the existence and uniqueness of the transport map, which is the manifold exponential of the gradient of a semi-convex map [41]. Displacement interpolation for transport on manifolds has the same variational characterization as the one introduced in [8] for Euclidean transport, see [55] for a detailed review of optimal transport on manifolds. Interpolation between pairs of measures generalizes to barycenters of a family of measures, see [37].

Displacement interpolation between Dirac measures amounts to computing a single geodesic curve on the manifold. Discretization and numerical solutions to this problem are numerous. A popular method is the Fast Marching algorithm introduced jointly by [49, 53] for isotropic Riemannian metrics (i.e. when the metric at each point is a scalar multiple of the identity) discretized on a rectangular grid. The complexity of the method is $O(N \log(N))$ operations, where

N is the number of grid points. This algorithm has been extended to compute geodesics on 2-D triangular meshes with only acute angles [38]. More general discretizations and the extension to Finsler metrics require the use of slower iterative schemes, see for instance [12].

Computing numerically optimal transport on manifolds has been less studied. For weighted sums of Diracs, displacement interpolation is achieved by solving the linear program to compute the coupling between the Diracs and then advancing the Diracs with the corresponding weights and constant velocity along the geodesics. In this article, we propose to extend the Eulerian discretization method [8] to solve for the displacement interpolation on a Riemannian manifold.

First order and proximal methods. The convex problem considered by Benamou and Brenier [7] can be re-casted as the optimization of a linear functional under second order conic constraints (see Section 3.6 for more details). This class of programs can be solved in time polynomial with the desired accuracy using interior points methods, see for instance [43].

However, the special structure of the problem, specially when discretized on an uniform grid, makes its suitable for first order scheme, and in particular proximal splitting methods. While they do not reach the same convergence speed for arbitrary conic program, they work well in practice for large scale problems, in particular when high accuracy is not mandatory, which is a common setup for problems in image processing. Proximal splitting schemes are first order optimization methods that allows one to minimize a sum of so-called "simple" functionals, possibly (for some methods) pre-composed by linear operators. A functional is called "simple" when it is possible to compute its proximal operator (see expression 1.1 for its precise definition) either in closed form, or with high accuracy using a few iterations of some sub-routine. In this article, we focus our attention to the Douglas-Rachford algorithm, introduced by [40] and on primal-dual methods. We make use of the recently proposed method [18], but other schemes could be used as well, see for instance [14]. We refer the reader to [21] and the references therein for more information about the properties of proximal maps and the associated proximal splitting schemes.

Note that the algorithm proposed by [7] corresponds to applying the Alternating Direction Method of Multiplier (ADDM) [30] to the Fenchel-Rockafeller dual of the (primal) dynamical transport problem. As shown by [28], this corresponds exactly to applying directly (a specific instanciation of) the Douglas-Rachford method to the primal problem.

Fluid mechanics discretization. While Lagrangian methods utilize a mesh-free discretizations (see for instance [36]), thats typically tracks the movement of centers of masses during the transportation, Eulerian method requires a fixed discretization of the spacial domain. The most straightforward strategy is to use an uniform centered discretization of an axis-aligned domain, which is used in most previously cited work, see for instance [7, 3]. Because of the close connection between dynamical optimal transport and fluid dynamics, we advocate in this article the use of staggered grids [2], which better copes with the incompressibility condition.

1.2. Contributions. Our first contribution is to show how the method initially proposed in [7] is a specific instance of the Douglas-Rachford algorithm. This allows one to use several variations on the initial method, by changing the values of the two relaxation parameters. Our second contribution is the introduction of a staggered grid discretization which is the canonical way to enforce incompressibility constraints. We show how this discretization fits into our proximal splitting methodology by introducing an interpolation operator and either making use of auxiliary variables or primal-dual methods. Our last contribution includes an exploration of several variations on the original convex transportation objective, the one proposed in [27], and a specially varying penalization which can be interpreted as replacing the L^2 ground cost by a geodesic distance on a Riemannian manifold. Note that the Matlab source code to reproduce the figures of this article is available online[1].

1.3. Notations. For a closed convex set $\mathcal{C}$, we denote its associated indicator function

$$\iota_{\mathcal{C}}(U) = \begin{cases} 0 & \text{if } U \in \mathcal{C}, \\ +\infty & \text{otherwise.} \end{cases}$$

[1] `http://www.ceremade.dauphine.fr/~peyre/codes/`

Given a convex, lower semi-continuous and proper functional $F : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ defined over some Hilbert space $\mathcal{H}$, its Legendre-Fenchel transform of F is defined as

$$F^*(V) = \max_{U \in \mathcal{H}} \langle U, V \rangle - F(U),$$

while its proximal operator $\mathrm{Prox}_{\gamma F} : \mathcal{H} \to \mathcal{H}$ is defined as

$$\mathrm{Prox}_{\gamma F}(U) = \operatorname*{argmin}_{U' \in \mathcal{H}} \frac{1}{2}\|U - U'\|^2 + \gamma F(U'). \tag{1.1}$$

Being able to compute this proximate mapping is crucial to use various proximal-splitting convex optimization methods. We call a function F such that $\mathrm{Prox}_{\gamma F}$ can be computed in closed form a *simple function*. Note that thanks to Moreau's identity

$$\mathrm{Prox}_{\gamma F^\star}(U) = U - \gamma\, \mathrm{Prox}_{F/\gamma}(U/\gamma), \tag{1.2}$$

computing the proximal operator of F^* has the same complexity as computing the proximal operator of F. We refer the reader to [21] for a detailed review of the various algebraic properties enjoyed by proximal maps.

2. Dynamical Optimal Transport Formulation.

2.1. Optimal Transport. In the following, we will restrict our exposition to maps $T : [0,1]^2 \mapsto [0,1]^2$ which satisfy homogeneous Dirichlet boundary conditions, and that are smooth. A valid transport map is a map that pushes forward the measure $f^0(x)\mathrm{d}x$ onto $f^1(x)\mathrm{d}x$. In term of densities, this corresponds to the constraint

$$f^0(x) = f^1(T(x))|\det(\partial T(x))|$$

where $\partial T(x) \in \mathbb{R}^{2\times 2}$ is the differential of T at x. This is known as the gradient equation. We call $\mathcal{T}(f^0, f^1)$ the set of transport that satisfies this constraint. An optimal transport T solves

$$\min_{T \in \mathcal{T}(f^0, f^1)} \int C(x, T(x))\mathrm{d}x \tag{2.1}$$

where $C(x,y) \geqslant 0$ is the cost of assigning $x \in [0,1]^2$ to $y \in [0,1]^2$. In the case $C(x,y) = \|x-y\|^2$ the optimal transport distance is called the L^2-Wasserstein distance between f^0 and f^1.

2.2. Fluid Mechanics Formulation. The geodesic path between the measures with densities $f^0(x)$ and $f^1(x)$ can be shown to have density $t \mapsto f(x,t)$ where $t \in [0,1]$ parameterize the path, where

$$f(x,t) = f^0((1-t)\mathrm{Id}_d + tT(x))|\det((1-t)\mathrm{Id}_d + t\partial T(x))|.$$

Benamou and Brenier showed in [7] that this geodesic solves the following non-convex problem over the densities $f(x,t) \in \mathbb{R}$ and a velocity field $v(x,t) \in \mathbb{R}^2$

$$\min_{(v,f) \in \mathcal{C}_v} \int_{[0,1]^d} \int_0^1 f(x,t)\|v(x,t)\|^2 \mathrm{d}t\mathrm{d}x, \tag{2.2}$$

under the set of non-linear constraints

$$\mathcal{C}_v = \left\{(v,f) \setminus \partial_t f + \mathrm{div}_x(fv) = 0,\ v(0,\cdot) = v(1,\cdot) = 0,\ f(\cdot,0) = f^0,\ f(\cdot,1) = f^1\right\}.$$

where the first relation in $\mathcal{C}_v$ is the continuity equation. We impose homogeneous Dirichlet conditions on the velocity field v and the temporal boundary constraints on f give the link with the density data.

Following [7], introducing the change of variable $(v,f) \mapsto (m,f)$, where m is the momentum $m = fv$, one obtains a convex optimization problem over the couple (f,m)

$$\min_{(m,f) \in \mathcal{C}} \mathcal{J}(m,f) = \int_{[0,1]^d} \int_0^1 J(m(x,t), f(x,t))\mathrm{d}t\mathrm{d}x, \tag{2.3}$$

$$\text{where} \quad \forall\,(m,f)\in\mathbb{R}^d\times\mathbb{R}, \quad J(m,f)=\begin{cases}\frac{\|m\|^2}{f} & \text{if } f>0,\\ 0 & \text{if } (m,f)=(0,0),\\ +\infty & \text{otherwise.}\end{cases} \tag{2.4}$$

and the set of linear constraints reads

$$\mathcal{C}=\left\{(m,f)\setminus \partial_t f+\mathrm{div}_x(m)=0,\ m(0,\cdot)=m(1,\cdot)=0,\ f(\cdot,0)=f^0,\ f(\cdot,1)=f^1\right\}.$$

3. Discretized Dynamic Optimal Transport. For simplicity of exposure, we describe the discretization for the 1-D case. It extends verbatim to higher dimensional discretization $d>1$.

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

$$\mathcal{G}_{\mathrm{c}}=\left\{(x_i=i/N,\, t_j=j/P)\in[0,1]^2\setminus 0\leqslant i\leqslant N, 0\leqslant j\leqslant P\right\}.$$

We denote

$$V=(m,f)\in\mathcal{E}_{\mathrm{c}}=(m_{i,j},f_{i,j}))_{0\leqslant i\leqslant N}^{0\leqslant j\leqslant P}$$

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

3.2. Staggered Grid. Similarly to the discretization of PDE's in incompressible fluid dynamics (see for instance [35]), we consider a staggered grid discretization which is more relevant to deal with the continuity equation:

$$\mathcal{G}_{\mathrm{s}}^x=\left\{\left(x_{i+1/2}=(i+1/2)/N,\, t_j=j/P\right)\in\frac{[-1,N+1]}{N}\times[0,1]\setminus -1\leqslant i\leqslant N, 0\leqslant j\leqslant P\right\},$$
$$\mathcal{G}_{\mathrm{s}}^t=\left\{\left(x_i=i/N,\, t_{j+1/2}=(j+1/2)/P\right)\in[0,1]\times\frac{[-1,P+1]}{P}\setminus 0\leqslant i\leqslant N, -1\leqslant j\leqslant P\right\}.$$

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

$$U=(\bar m,\bar f)\in\mathcal{E}_{\mathrm{s}}=((\bar m_{i,j})_{-1\leqslant i\leqslant N}^{0\leqslant j\leqslant P},(\bar f_{i,j})_{0\leqslant i\leqslant N}^{-1\leqslant j\leqslant P})$$

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

3.3. Interpolation and Divergence Operators. We introduce a midpoint interpolation operator $\mathcal{I}:\mathcal{E}_{\mathrm{s}}\to\mathcal{E}_{\mathrm{c}}$, where, for $U=(\bar m,\bar f)\in\mathcal{E}_{\mathrm{s}}$, we define $\mathcal{I}(U)=(m,f)\in\mathcal{E}_{\mathrm{c}}$ as

$$\forall\, 0\leqslant i\leqslant N,\quad \forall\, 0\leqslant j\leqslant P,\quad \begin{cases} m_{i,j} &= (\bar m_{i+1/2,j}+\bar m_{i-1/2,j})/2,\\ f_{i,j} &= (\bar f_{i,j+1/2}+\bar f_{i,j-1/2})/2.\end{cases} \tag{3.1}$$

The space-time divergence operator $\mathrm{div}:\mathcal{E}_{\mathrm{s}}\to\mathbb{R}^{\mathcal{G}_{\mathrm{c}}}$ is defined, for $U=(\bar m,\bar f)\in\mathcal{E}_{\mathrm{s}}$ as

$$\forall\, 0\leqslant i\leqslant N,\ \forall\, 0\leqslant j\leqslant P,\quad \mathrm{div}(U)_{i,j}=(\bar m_{i+1/2,j}-\bar m_{i-1/2,j})+(\bar f_{i,j+1/2}-\bar f_{i,j-1/2}).$$

3.4. Boundary Constraints. We extract the boundary values on the staggered grid using the linear operator b, defined, for $U=(\bar m,\bar f)\in\mathcal{E}_{\mathrm{s}}$ as

$$b(U)=\left((\bar m_{-\frac12,j},\bar m_{N+\frac12,j})_{j=0}^P,(\bar f_{i,-1/2},\bar f_{i,P+1/2})_{i=0}^N\right)\in\mathbb{R}^{P+1}\times\mathbb{R}^{P+1}\times\mathbb{R}^{N+1}\times\mathbb{R}^{N+1}.$$

We impose the following boundary values

$$b(U)=b_0\quad\text{where}\quad b_0=(0,0,f^0,f^1)\in\mathbb{R}^{P+1}\times\mathbb{R}^{P+1}\times\mathbb{R}^{N+1}\times\mathbb{R}^{N+1},$$

where $f^0,f^1\in\mathbb{R}^{N+1}$ are the discretized initial (time $t=0$) and final (time $t=1$) densities and the spatial boundary constraint $0\in\mathbb{R}^P$ on the momentum $m=fv$ comes from the discretized Dirichlet boundary conditions on the velocity field v.

3.5. Discrete Convex Problem. The initial continuous problem (2.3) is approximated on the discretization grid by solving the finite dimensional convex problem

$$\min_{U \in \mathcal{E}_{\mathrm{s}}} \mathcal{J}(\mathcal{I}(U)) + \iota_{\mathcal{C}}(U). \tag{3.2}$$

The discrete objective functional $\mathcal{J}$ reads for $V = (m, f) \in \mathcal{E}_{\mathrm{c}}$:

$$\mathcal{J}(V) = \sum_{k \in \mathcal{G}_{\mathrm{c}}} J(m_k, f_k), \tag{3.3}$$

where we denote $k = (i, j) \in \mathcal{G}_{\mathrm{c}}$ the indexes on the centered grid, and the functional J is defined in (2.4).

The constraint set $\mathcal{C}$ corresponds to the divergence-free constraint together with the boundary constraints

$$\mathcal{C} = \{U \in \mathcal{G}_{\mathrm{s}} \setminus \operatorname{div}(U) = 0 \quad \text{and} \quad b(U) = b_0\}.$$

3.6. Second Order Cone Programming Formulation. The discretized problem (3.2) can be equivalently solved as the following minimization over variables $(U, V, W) \in \mathcal{E}_{\mathrm{s}} \times \mathcal{E}_{\mathrm{c}} \times \mathbb{R}^{\mathcal{G}_{\mathrm{c}}}$

$$\min_{(V,W) \in \mathcal{K}, V = \mathcal{I}(U)} \sum_{k \in \mathcal{G}_{\mathrm{c}}} W_k \tag{3.4}$$

where $\mathcal{K}$ is the product of (rotated) Lorentz cones

$$\mathcal{K} = \left\{(V = (\bar{m}, \bar{f})), W) \in \mathcal{E}_{\mathrm{c}} \times \mathbb{R}^{\mathcal{G}_{\mathrm{c}}} \setminus \forall\, k \in \mathcal{G}_{\mathrm{c}},\ \|\bar{m}_k\|^2 - W_k \bar{f}_k \leqslant 0\right\}.$$

Problem (3.4) is a specific instance of so-called second-order cone program (SOCP), that can be solved in time polynomial with the accuracy using interior point methods (see Section 1.1 for more details). As already mentioned in Section 1.1, we focus in this article on proximal splitting methods, that are more adapted to large scale imaging problems.

4. Proximal Splitting Algorithms. In this section, we first compute the proximal operators of the cost function J and of indicator of the constraint set $\mathcal{C}$. We then present some proximal splitting algorithms that can be used to solve (3.2).

4.1. Computing $\operatorname{Prox}_{\gamma \mathcal{J}}$. The following proposition shows that the functional $\mathcal{J}$ defined in (3.3) is simple, in the sense that its proximal operator can be computed in closed form.

Proposition 1. *One has*

$$\forall\, V \in \mathcal{E}_c, \quad \operatorname{Prox}_{\gamma \mathcal{J}}(V) = (\operatorname{Prox}_{\gamma J}(V_k))_{k \in \mathcal{G}_c}$$

where, for all $(\tilde{m}, \tilde{f}) \in \mathbb{R}^d \times \mathbb{R}$,

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

$$\text{where} \quad \forall\, f \geqslant 0, \quad \mu(f) = \frac{f \tilde{m}}{f + 2\gamma} \tag{4.1}$$

and $f^\star$ is the largest real root of the third order polynomial equation in X

$$P(X) = (X - \tilde{f})(X + 2\gamma)^2 - \gamma \|\tilde{m}\|^2 = 0. \tag{4.2}$$

Proof. This proposition is a special case of Proposition 2. □

Note that since J is a positive 1-homogenous functional, its Legendre-Fenchel transform is the indicator of a convex set

$$J^* = \iota_{\mathcal{C}_J} \quad \text{where} \quad \mathcal{C}_J = \left\{(a, b) \in \mathbb{R}^d \times \mathbb{R} \setminus 2\|a\|^2 + b \leqslant 0\right\}.$$

This allows to re-write the functional J as

$$J(m,f) = \sup_{(a,b)\in\mathcal{C}_J} \langle a,\, m\rangle + b\, f$$

which is essentially the approach used in [7] to apply the ADMM algorithm on a dual problem. Note that to relate this approach to the Douglas-Rachford algorithm described bellow, one can directly use the Moreau's identity (1.2) to compute the proximal operator of J using the orthogonal projection on $\mathcal{C}_j$

$$\mathrm{Prox}_{\gamma J}(v) = v - \gamma\, \mathrm{Proj}_{\mathcal{C}_J}(v/\gamma).$$

4.2. Computing $\mathrm{Proj}_{\mathcal{C}}$. The proximal mapping of $\iota_{\mathcal{C}}$ is $\mathrm{Proj}_{\mathcal{C}}$ the orthogonal projector on the convex set $\mathcal{C}$. This is an affine set that can be written as

$$\mathcal{C} = \{U = (m,f) \in \mathcal{E}_{\mathrm{s}} \setminus AU = y\} \quad \text{where} \quad AU = (\mathrm{div}(U), b(U)) \quad \text{and} \quad y = (0, b_0).$$

This projection can be computed by solving a linear system as

$$\mathrm{Proj}_{\mathcal{C}} = (\mathrm{Id} - A^*\Delta^{-1}A) + A^*\Delta^{-1}y$$

where applying $\Delta^{-1} = (AA^*)^{-1}$ requires solving a Poisson equation on the centered grid $\mathcal{G}_{\mathrm{c}}$ with the prescribed boundary condition. 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 [51].

4.3. Douglas-Rachford Solver. The Douglas-Rachford (DR) algorithm [40] is a proximal splitting method that allows one to minimize the sum of two simple functions.

Splitting reformulation.. We recast the initial optimal transport problem (3.2) as

$$\min_{z=(U,V)\in\mathcal{E}_{\mathrm{s}}\times\mathcal{E}_{\mathrm{c}}} G_1(z) + G_2(z) \tag{4.3}$$

where we have defined the functionals

$$\forall\, z = (U,V) \in \mathcal{E}_{\mathrm{s}} \times \mathcal{E}_{\mathrm{c}}, \quad G_1(z) = \mathcal{J}(V) + \iota_{\mathcal{C}}(U) \quad \text{and} \quad G_2(z) = \iota_{\mathcal{C}_{s,c}}(z).$$

In this expression, $\mathcal{C}_{s,c}$ is the constraint set that couples staggered and centered variables

$$\mathcal{C}_{s,c} = \{z = (U,V) \in \mathcal{E}_{\mathrm{s}} \times \mathcal{E}_{\mathrm{c}} \setminus V = \mathcal{I}(U)\}$$

and $\mathcal{I}$ is the interpolation operator defined in (3.1).

Proximal operators for DR.. Both G_1 and G_2 are simple since

$$\begin{aligned}
&\mathrm{Prox}_{\gamma G_1}(U,V) = (\mathrm{Proj}_{\mathcal{C}}(U), \mathrm{Prox}_{\gamma\mathcal{J}}(V)),\\
&\mathrm{Prox}_{\gamma G_2}(U,V) = \mathrm{Proj}_{\mathcal{C}_{s,c}}(U,V) = (\tilde{U}, \mathcal{I}(\tilde{U})) \quad \text{where} \quad \tilde{U} = (\mathrm{Id} + \mathcal{I}^*\mathcal{I})^{-1}(U + \mathcal{I}^*(V))
\end{aligned}$$

where $\mathcal{I}^*$ is the adjoint of the linear interpolation operator. Note that computing $\mathrm{Proj}_{\mathcal{C}_{s,c}}$ 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.

DR iterations.. The iterates of the DR algorithm defines a sequence $(z^{(\ell)}, w^{(\ell)}) \in (\mathcal{E}_{\mathrm{s}} \times \mathcal{E}_{\mathrm{c}}) \times (\mathcal{E}_{\mathrm{s}} \times \mathcal{E}_{\mathrm{c}})$ using a initial $w^{(0)}$ and

$$z^{(\ell)} = \mathrm{Prox}_{\gamma G_2}(w^{(\ell)}), \tag{4.4}$$

$$w^{(\ell+1)} = \left(1 - \frac{\alpha}{2}\right) w^{(\ell)} + \frac{\alpha}{2}\, \mathrm{RProx}_{\gamma G_2} \circ \mathrm{RProx}_{\gamma G_1}(w^{(\ell)}) \tag{4.5}$$

and where we used the following shortcut notation for the reflexive proximal mapping

$$\mathrm{RProx}_{\gamma G}(z) = 2\,\mathrm{Prox}_{\gamma G}(z) - z.$$

If $0 < \alpha < 2$ and $\gamma > 0$, one can show that $z^{(\ell)} \to z^\star$ a solution of (4.3), see [20] for more details. In our case, the iterates reads $z^{(\ell)} = (U^{(\ell)}, V^{(\ell)})$, which allows one to retrieve an approximation $f^{(\ell)}$ of the transport geodesic as $U^{(\ell)} = (m^{(\ell)}, f^{(\ell)})$.

In the case that $\alpha = 1$ (no relaxation), one can show that the DR algorithm is equivalent to the ADMM on the dual [28] and to the ALG2 [30] used by Benamou and Brenier in [7] to solve the optimal transport problem on a centered grid. Since we consider a staggered grid, the use of the interpolation operator makes our optimization problem (4.3) different from the original ALG2 and requires the introduction of an auxiliary variable V. Furthermore, the introduction of an extra relaxation parameter α is useful to speed-up the convergence of the method, as will be established in the experimentations.

The iteration (4.4) guarantees that the iterates satisfy $U^{(\ell)} = \mathcal{I}(V^{(\ell)})$. It is also possible to exchange the roles of G_1 and G_2, which defines a different algorithm, for which the iterates satisfy $\mathrm{div}(U^{(\ell)}) = 0$ and $V^{(\ell)}$ is a positive density (but condition $U^{(\ell)} = \mathcal{I}(V^{(\ell)})$ does not hold in general).

4.4. Primal-Dual Solver. Primal dual (PD) algorithm such as the relaxed Arrow-Hurwitz method introduced in [18] allows one to minimize functionals of the form $G_1 + G_2 \circ A$ where A is a linear operator and G_1, G_2 are simple functions. One can thus directly apply this method to problem (3.2) with $G_2 = \mathcal{J}$, $A = \mathcal{I}$ and $G_1 = \iota_{\mathcal{C}}$.

The iterations compute a sequence $(U^{(\ell)}, \Upsilon^{(\ell)}, V^{(\ell)}) \in \mathcal{E}_{\mathrm{s}} \times \mathcal{E}_{\mathrm{s}} \times \mathcal{E}_{\mathrm{c}}$ of variables from an initial $(\Upsilon^{(0)}, V^{(0)})$ according to

$$\begin{aligned} V^{(\ell+1)} &= \mathrm{Prox}_{\sigma G_2^*}(V^{(\ell)} + \sigma \mathcal{I} \Upsilon^{(\ell)}), \\ U^{(\ell+1)} &= \mathrm{Prox}_{\tau G_1}(U^{(\ell)} - \tau \mathcal{I}^* V^{(\ell+1)}), \\ \Upsilon^{(\ell+1)} &= U^{(\ell+1)} + \theta(U^{(\ell+1)} - U^{(\ell)}). \end{aligned}$$

Note that $\mathrm{Prox}_{\sigma G_2^*}$ can be computed using Prox_{G_2} following equation (1.2). If $0 \leqslant \theta \leqslant 1$ and $\sigma\tau\|\mathcal{I}\|^2 < 1$ then one can prove that $U^{(\ell)} \to U^\star$ which is a solution of (3.2).

The case $\theta = 0$ corresponds to the Arrow-Hurwicz algorithm [4], and the general case can be interpreted as a preconditioned version of the ADDM algorithm.

5. Generalized Cost Functions. Following [27, 17], we propose to use a generalized cost function that allows one to compute geodesics that interpolate between the L^2-Wasserstein and the H^{-1} geodesics. To introduce further flexibility, we introduce a spacial varying weight.

5.1. Spatially Varying Interpolation between L^2-Wasserstein and H^{-1} . We define our new general finite dimensional convex problem as

$$\min_{U \in \mathcal{E}_{\mathrm{s}}} \mathcal{J}_\beta^w(\mathcal{I}(U)) + \iota_{\mathcal{C}}(U), \tag{5.1}$$

where the vector of weights $w = (w_k)_{k \in \mathcal{G}_{\mathrm{c}}}$ satisfies $c < w_k \leqslant +\infty$ and $c > 0$ is a small constant. The generalized functional reads, for $\beta \geqslant 0$

$$\mathcal{J}_\beta^w(V) = \sum_{k \in \mathcal{G}_{\mathrm{c}}} w_k J_\beta(m_k, f_k), \tag{5.2}$$

$$\forall\, (m, f) \in \mathbb{R}^d \times \mathbb{R}, \quad J_\beta(m, f) = \begin{cases} \frac{\|m\|^2}{f^\beta} & \text{if } f > 0, \\ 0 & \text{if } (m, f) = (0, 0), \\ +\infty & \text{otherwise.} \end{cases} \tag{5.3}$$

Note that $\mathcal{J}_1^1 = \mathcal{J}$ and that for $\beta \in [0, 1]$, $\mathcal{J}_\beta^w$ is convex. Furthermore, one has, for $f > 0$

$$\det(\partial^2 J_\beta(m, f)) = \frac{4\beta(1-\beta)\|m\|^2}{f^{3\beta+2}}. \tag{5.4}$$

This shows that J_β is strictly convex on $\mathbb{R}^{*,d} \times \mathbb{R}^{+,*}$ for $\beta \in]0, 1[$.

When $\beta = 1$ and the weights w_k are constant in time, the solution of (5.2) discretizes the displacement interpolation between the densities (f^0, f^1) for a ground cost being the squared

geodesic distance on a Riemannian manifold (see Section 1.1 for more details). Note that we restrict our attention to isotropic Riemannian metrics (being propositional to the identity at each point), but this extends to arbitrary Riemannian or even Finsler metrics. Studying the properties of this transportation distance is however not the purpose of this work, and we show in Section 6.3 numerical illustrations of the influence of w.

5.2. Computing $\mathrm{Prox}_{\gamma J_\beta}$**.** The following proposition, which generalizes Proposition 1, shows how to compute $\mathrm{Prox}_{\gamma \mathcal{J}_\beta^w}$.

PROPOSITION 2. *One has*

$$\forall\, V \in \mathcal{E}_c, \quad \mathrm{Prox}_{\gamma \mathcal{J}_\beta^w}(V) = (\mathrm{Prox}_{\gamma w_k J_\beta}(V_k))_{k \in \mathcal{G}_c}$$

where, for all $(\tilde{m}, \tilde{f}) \in \mathbb{R}^d \times \mathbb{R}$,

$$\mathrm{Prox}_{\gamma J_\beta}(\tilde{m}, \tilde{f}) = \begin{cases} (\mu(f^\star), f^\star) & \text{if } f^\star > 0 \text{ and } \gamma < \infty, \\ (0,0) & \text{otherwise.} \end{cases}$$

$$\text{where} \quad \forall\, f \geqslant 0, \quad \mu(f) = \frac{f^\beta \tilde{m}}{f^\beta + 2\gamma} \in \mathbb{R}^d \tag{5.5}$$

and $f^\star$ *is the largest real root of the following equation in* X

$$P_\beta(X) = X^{1-\beta}(X - \tilde{f})(X^\beta + 2\gamma)^2 - \gamma\beta\|\tilde{m}\|^2 = 0 \tag{5.6}$$

Proof. We denote $(\bar{m}, \bar{f}) = \mathrm{Prox}_{\gamma J_\beta}(\tilde{m}, \tilde{f})$, and

$$\forall\, (m, f) \in \mathbb{R}^d \times \mathbb{R}, \quad \Phi_\beta(m, f) = \frac{1}{2}\|(m, f) - (\tilde{m}, \tilde{f})\|^2 + \gamma J_\beta(m, f).$$

If $\bar{f} > 0$, since J is C^1 and is strongly convex on $\mathbb{R}^d \times \mathbb{R}^{+,*}$, necessarily $(\bar{m}, \bar{f})$ is the unique solution of $\nabla \Phi_\beta(\bar{m}, \bar{f}) = 0$, which reads

$$\begin{cases} 2\gamma \frac{\bar{m}}{\bar{f}^\beta} + \bar{m} - \tilde{m} = 0 \\ -\gamma\beta \frac{\|\bar{m}\|^2}{\bar{f}^{\beta+1}} + \bar{f} - \tilde{f} = 0. \end{cases}$$

Reformulating these equations leads to the following equivalent conditions

$$P_\beta(\bar{f}) = 0 \quad \text{and} \quad \bar{m} = \mu_\beta(\bar{f}).$$

This shows that if P_β as at least a strictly positive real root $f^\star$, it is necessarily unique and that $(\bar{f} = f^\star, \bar{m} = \mu_\beta(f^\star))$. Otherwise, necessarily $(\bar{f} = 0, \bar{m} = 0)$. □

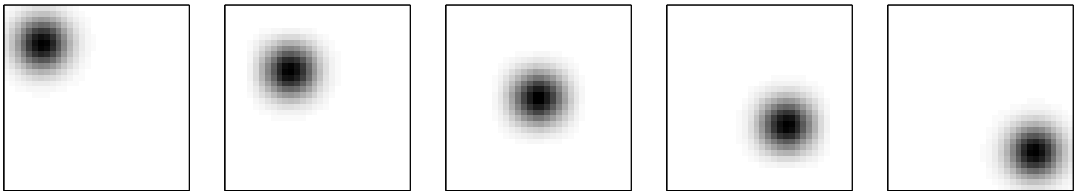
Note that when $\beta = p/q \in \mathbb{Q}$ is a rational number, equation $P_\beta(X) = 0$ corresponds to finding the root of a polynomial. It can be solved efficiently using a few Newton descent iterations starting from a large enough value of f.

6. Numerical Simulations.

6.1. Comparison of Proximal Schemes. This Section compares the following algorithms introduced in Section 4:

- Douglas-Rachford (DR in the following) as exposed in Section 4.3, parameterized with α and γ ;
- ADDM on the dual (ADDM in the following) introduced in [7], (which is DR with $\alpha = 1$) parameterized with γ ;
- Primal-dual (PD in the following) as exposed in Section 4.4, parameterized with σ and τ.

The comparison is done on a simple example with two 2-D isotropic Gaussian distributions (f^0, f^1) with the same variance. In the continuous case, the solution is known to be a translation between the mean of the Gaussians. The spatial domain is here of dimension $d = 2$ and it is discretized on a grid with $(N = P = 31)$ points for both each dimension. We compute an (almost)



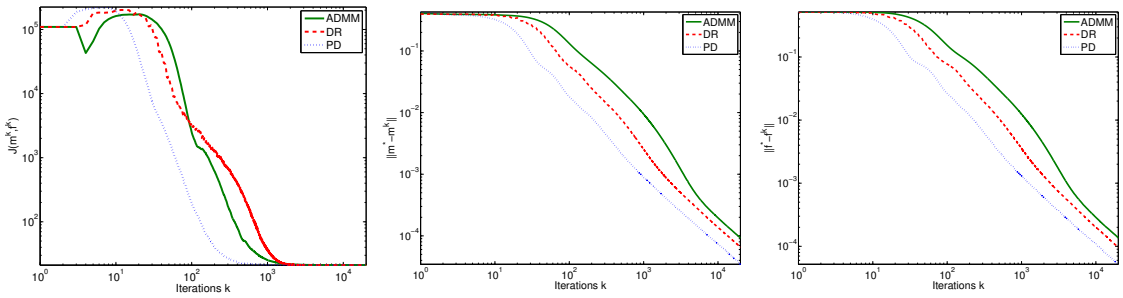
$t = 0$ $t = 1/4$ $t = 1/2$ $t = 3/4$ $t = 1$

FIG. 6.1. *Display of $f^\star(\cdot, t)$ for several value of t (note that for $t = 0$ and $t = 1$, this corresponds to f^0 and f^1). The grayscale values are linearly interpolating from black to white between 0 and the maximum value of $f^\star$.*

exact reference solution $(m^\star, f^\star)$ of the discrete problem with 10^6 iterations of the DR. The obtained transported mass $f^\star(\cdot, t)$ is illustrated in Figure 6.1.

For each algorithm, we perform an exhaustive search of the best possible set of parameters. These optimal parameters are those minimizing $\|(m^\star, f^\star) - (m^{(\ell)}, f^{(\ell)})\|$, the ℓ^2 distance between $f^\star$ and the output of the algorithm after $\ell = 100$ iterations. The obtained parameters for this data set are: $\gamma = 1/350$ for ADMM on the dual, $(\gamma = 1/240, \alpha = 1.983)$ for DR and $\sigma = 85$ for PD. For PD, we found that simply setting $\tau = \frac{0.99}{\sigma\|\mathcal{I}\|^2}$ leads to almost optimal convergence rate in our tests, so we use this rule to only introduce a single parameter σ. Note that this parameter choice is within the range of parameters $\sigma\tau\|\mathcal{I}\|^2 < 1$ that guaranties convergence of the PD method. Figure 6.2 displays, for this optimal choice of parameters, the evolution of the cost function value as well as the convergence toward $(m^\star, f^\star)$ as a function of the iteration number ℓ.

One can observe that the quality of the estimation can not be deduced from the cost function evolution since the functional is very flat. Indeed, an almost minimal value of the function minimum is reached by all the algorithms after roughly 10^3 iterations, whereas the ℓ^2 distance to the reference solution continue to decrease almost linearly in log-log scale.



$\mathcal{J}(m^{(\ell)}, f^{(\ell)})$ $\|m^\star - m^{(\ell)}\|$ $\|f^\star - f^{(\ell)}\|$

FIG. 6.2. *At each iteration ℓ, we plot the value of the cost function $\mathcal{J}$ and the distance between the reference solution $(m^\star, f^\star)$ and the estimation $(m^{(\ell)}, f^{(\ell)})$ for the different proximal splitting algorithms with the best found parameters.*

When comparing the three approaches, PD presents the fastest convergence rate to the reference solution and the relaxed DR algorithm performs better than ADDM on the dual, which shows the advantage of introducing the α parameter. Note also that the computational cost is smaller for PD, as it takes[2] 0.13s for one PD iteration and $0.2s$ for one DR or ADMM iteration for this example.

6.2. Interpolation Between L^2-Wasserstein and H^{-1}. We first apply the PD algorithm for different values of β on the bump example introduced in the previous section. The results are presented in the Figure 6.3, which shows the level-lines of the estimated densities $f^{(\ell)}(\cdot, t)$ for $\ell = 1000$ iterations. It shows the evolution of the solution between a H^{-1} linear interpolation $(\beta = 0)$ and a displacement interpolation with transport $(\beta = 1)$.

Figure 6.4 shows for $\beta = 1/2$ and $\beta = 3/4$ the evolution of the cost function with the iterations index ℓ, together with the convergence of the estimate to the reference solution $(m^\star, f^\star)$ (obtained after 10^5 iterations of the PD algorithm). We can observe that the behavior of the process with

[2]For our Matlab implementation on a standard laptop computer.

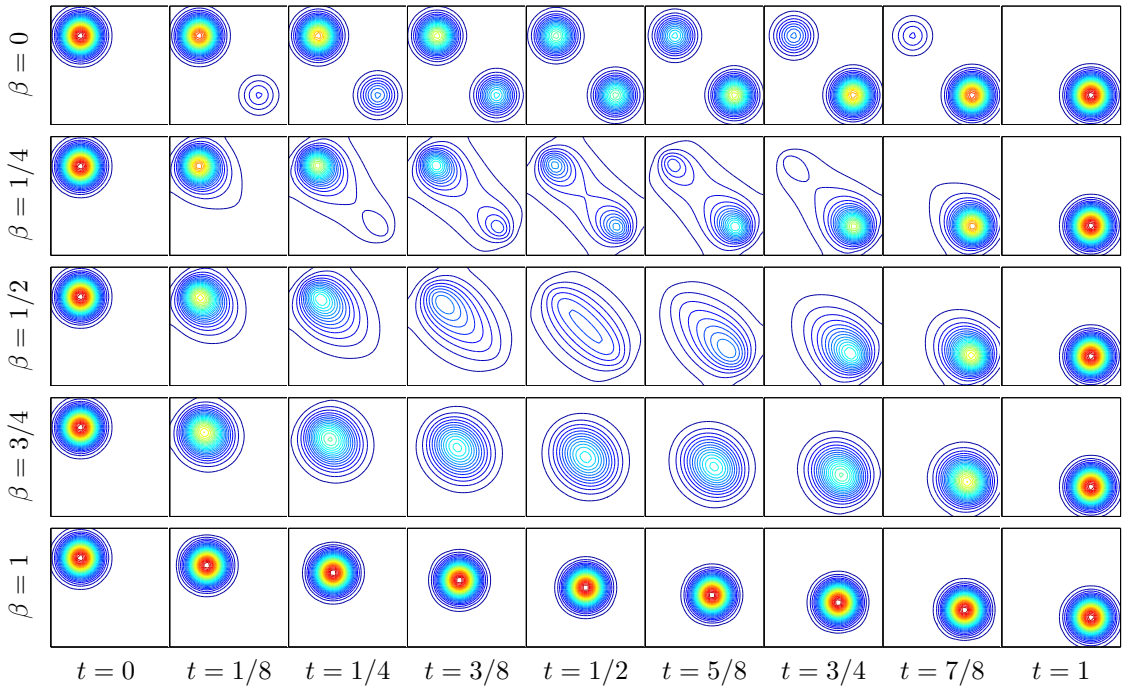


FIG. 6.3. *Display of the level sets of $f^{(\ell)}(\cdot, t)$ for several value of t and β (note that for $t = 0$ and $t = 1$, this corresponds to f^0 and f^1).*

$\beta \in]0, 1[$ is different than the one observed for $\beta = 1$ in Figure 6.2. Indeed, we observe a faster convergence of the algorithms, which is consistent with the fact that $\mathcal{J}_\beta$ becomes more and more strongly convex as β approaches $1/2$ (see (5.4)).

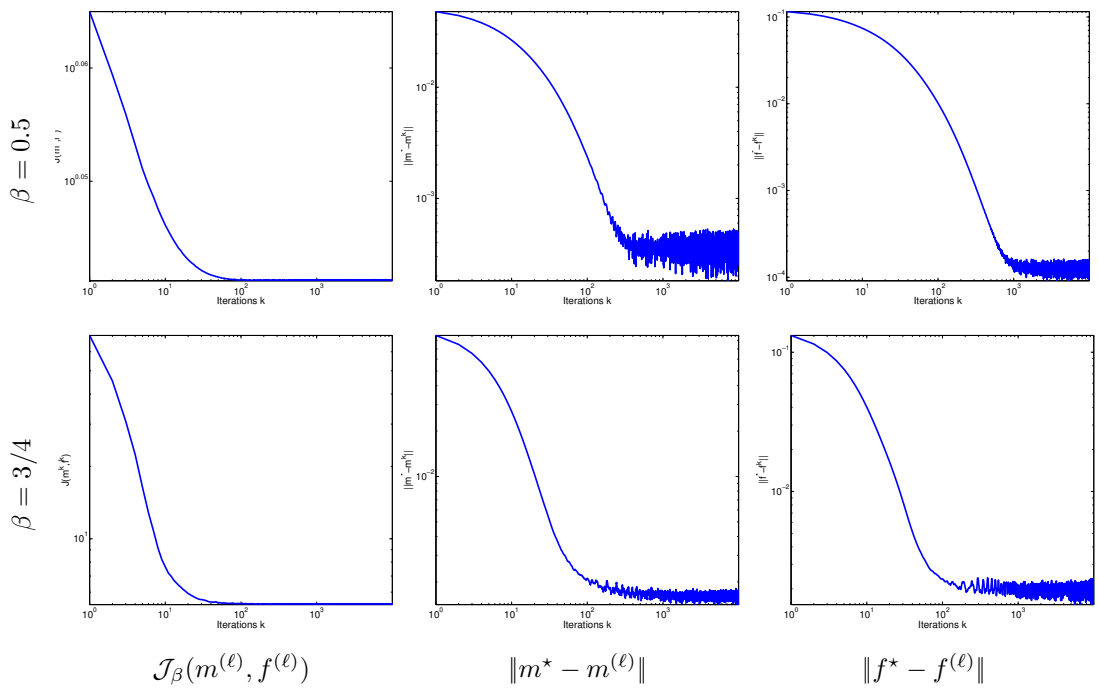


FIG. 6.4. *At each iteration ℓ, we plot the value of the cost function $\mathcal{J}_\beta(m^{(\ell)}, f^{(\ell)})$ and the distance between the reference solution $(m^\star, f^\star)$ and the estimation $(m^{(\ell)}, f^{(\ell)})$. The first (resp. second) row presents the result with $\beta = 1/2$ (resp. $\beta = 3/4$).*

As a second example, we show in Figure 6.5 the different morphings obtained between pictures of Gaspard Monge and Leonid Kantorovich. The grayscale representation scales linearly between black (value of 0) and white (value of 1) and the dimensions are $N + 1 = 75$, $M + 1 = 100$,

$P+1=60$, M being the number of discrete points in the second spatial dimension.

FIG. 6.5. *Evolution of $f^{(\ell)}(\cdot,t)$ for several value of t and β. The first and last columns represent the data f^0 and f^1. The intermediate ones present the reference solution $f^*(t)$ for successive times $t = i/6$, $i = 1\cdots 5$. Each line illustrates f^* for different values $\beta = j/4$, $j = 0\cdots 4$ of the generalized cost function.*

6.3. Riemannian Transportation. We investigate in this section the approximation of a displacement interpolation for a ground cost being the squared geodesic distance on a Riemannian manifold. This is achieved by solving (5.1) with $\beta = 1$ but a non-constant weight map w.

We exemplify this setting by considering optimal transport with obstacles, which corresponds to choosing weights w that are infinity on the obstacle $\mathcal{O} \subset \mathbb{R}^d \times \mathbb{R}$, i.e.

$$\forall\, k \in \mathcal{G}_{\mathrm{c}}, \quad w_k = 1 + \iota_{\mathcal{C}}(x_k, t_k) \in \{1, +\infty\}.$$

Note that the obstacles can be dynamic, i.e. the weight w needs not to be constant in time.

Figure 6.6 shows a first example where $\mathcal{O}$ is a 2-D ($d = 2$) static labyrinth map (the walls of the labyrinth being the obstacles, and are displayed in black). We use a $50 \times 50 \times 100$ discretization grid of the space-time domain $[0,1]^3$ and the input measures (f^0, f^1) are Gaussians with standard deviations equal to 0.04. For Gaussians with such a small variance, this example shows that the displacement interpolation is located closely to the geodesic path between the centers of the gaussians.

Figure 6.7 shows a more complicated setting that includes a labyrinth with moving walls: a wall appears at time $t = 1/4$ and another one disappears at time $1/2$. The difference with respect to the

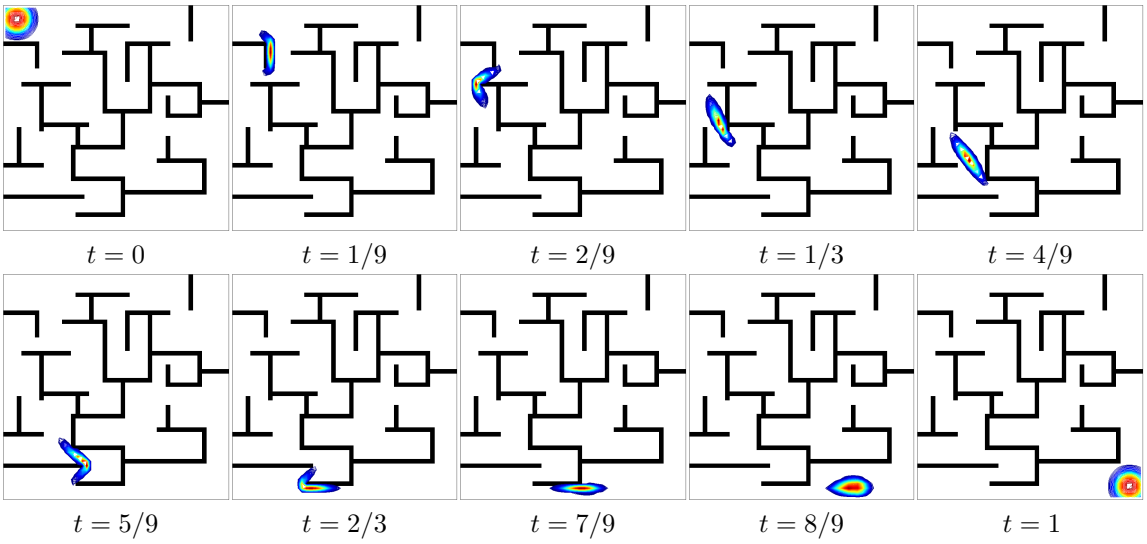


FIG. 6.6. *Evolution of $f^{(\ell)}(\cdot, t)$ for several values of t, using a Riemannian manifold with weights w_k (constant in time) restricting the densities to lie within a 2-D static labyrinth map.*

previous example is the fact that w is now time dependent. This simple modification has a strong impact on the displacement interpolation. Indeed, the speed of propagation of the mean of the density is not constant anymore since the density measure is confined in a small area surrounded by walls for $t \in [1/4, 1/2]$.

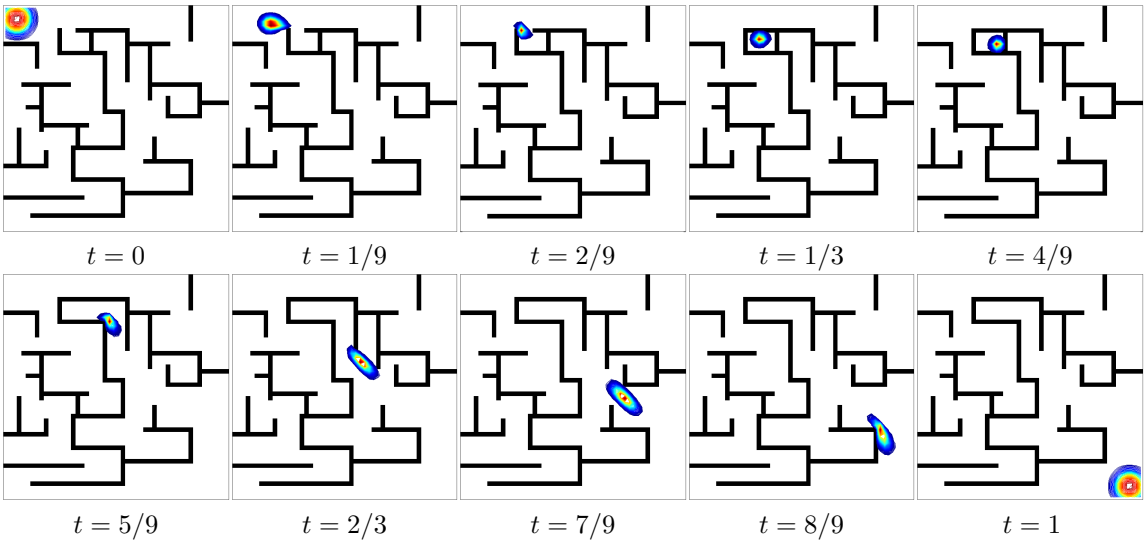


FIG. 6.7. *Evolution of $f^{(\ell)}(\cdot, t)$ for several values of t, using a Riemannian manifold with weights w_k (evolving in time) restricting the densities to lie within a 2-D dynamic labyrinth map (i.e. with moving walls).*

Conclusion. In this article, we have shown how proximal splitting schemes offer an elegant and unifying framework to describe computational methods to solve the dynamical optimal transport with an Eulerian discretization. This allowed use to extend the original method of Benamou and Brenier in several directions, most notably the use of staggered grid discretization and the introduction of generalized, spatially variant, cost functions.

Acknowledgment. We would like to thank to Jalal Fadili for his detailed explanations of the connexion between the ADMM and DR algorithms. Gabriel Peyré acknowledges support from the European Research Council (ERC project SIGMA-Vision). This work was also partially funded by the French Agence Nationale de la Recherche (ANR) under reference ANR-11-BS01-014-01.

REFERENCES

[1] M. Agueh and G. Carlier. Barycenters in the wasserstein space. SIAM Journal on Mathematical Analysis, 43(2):904–924, 2011.

[2] J. Anderson. Computational Fluid Dynamics. McGraw-Hill Science/Engineering/ Math, 1995.

[3] S. Angenent, S. Haker, and A. Tannenbaum. Minimizing flows for the monge-kantorovich problem. SIAM Journal on Mathematical Analysis, 35(1):61–97, 2003.

[4] K.J. Arrow, L. Hurwicz, and H. Uzawa. Studies in linear and non-linear programming. Stanford University Press, 1958.

[5] J.-D. Benamou. A domain decomposition method for the polar factorization of vector-valued mappings. SIAM Journal on Numerical Analysis, 32(6):1808–1838, 1995.

[6] J.-D. Benamou. Numerical resolution of an "unbalanced" mass transport problem. ESAIM: Mathematical Modelling and Numerical Analysis, 37(5):851–868, 2010.

[7] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution of the monge-kantorovich mass transfer problem. Numerische Mathematik, 84(3):375–393, 2000.

[8] J.-D. Benamou and Y. Brenier. Wasserstein optimal mapping between prescribed densities functions. Journal of Optimization Theory and Applications, 111(2):255–271, 2001.

[9] J.-D. Benamou, B. D. Froese, and Oberman A. M. Numerical solution of the optimal transportation problem via viscosity solutions for the monge-ampere equation. CoRR, abs/1208.4873, 2012.

[10] D.P. Bertsekas. The auction algorithm: A distributed relaxation method for the assignment problem. Annals of Operations Research, 14:105–123, 1988.

[11] N. Bonneel, M. van de Panne, S. Paris, and W. Heidrich. Displacement interpolation using lagrangian mass transport. ACM Transactions on Graphics (SIGGRAPH ASIA'11), 30(6), 2011.

[12] F. Bornemann and C. Rasch. Finite-element discretization of static hamilton-jacobi equations based on a local variational principle. Comput. Visual Sci., 9(2):57–69, 2006.

[13] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. Communications on Pure and Applied Mathematics, 44(4):375–417, 1991.

[14] L. M. Briceño-Arias and P. L. Combettes. A monotone+skew splitting model for composite monotone inclusions in duality. SIAM Journal on Optimization, 21(4):1230–1250, 2011.

[15] M. Burger, M. Franek, and C.-B. Schölieb. Regularized regression and density estimation based on optimal transport. Applied Mathematics Research eXpress, 2011.

[16] R. Burkard, M. Dell'Amico, and S. Martello. Assignment Problems. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2009.

[17] P. Cardaliaguet, G. Carlier, and B. Nazaret. Geodesics for a class of distances in the space of probability measures. Calculus of Variations and Partial Differential Equations, pages 1–26, 2012.

[18] 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(1):120–145, 2011.

[19] P. Clarysse, B. Delhay, M. Picq, and J. Pousin. Optimal extended optical flow subject to a statistical constraint. Journal of Computational and Applied Mathematics, 234(4):1291–1302, 2010.

[20] P.L. Combettes and J.-C. Pesquet. A douglas-rachford splitting approach to nonsmooth convex variational signal recovery. IEEE Journal of Selected Topics in Signal Processing, 1(4):564 –574, 2007.

[21] P.L. Combettes and J-C. Pesquet. Proximal splitting methods in signal processing. In Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pages 185–212. Springer, 2011.

[22] G. B. Dantzig. Linear Programming and Extensions. Princeton University Press, Princeton, NJ, 1963.

[23] F. de Goes, D. Cohen-Steiner, P. Alliez, and M. Desbrun. An optimal transport approach to robust reconstruction and simplification of 2d shapes. Computer Graphics Forum, 30(5):1593–1602, 2011.

[24] 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., 22:71–96, 2006.

[25] J. Delon, J. Salomon, and A. Sobolevski. Fast Transport Optimization for Monge Costs on the Circle. SIAM Journal on Applied Mathematics, 70(7):2239–2258, 2010.

[26] J. Delon, J. Salomon, and A. Sobolevskii. Local matching indicators for transport problems with concave costs. SIAM Journal on Discrete Mathematics, 26(2):801–827, 2012.

[27] J. Dolbeault, B. Nazaret, and G. Savaré. A new class of transport distances between measures. Calculus of Variations and Partial Differential Equations, 34(2):193–231, 2009.

[28] J. Eckstein and D. P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming, 55:293–318, 1992.

[29] X. Feng and M. Neilan. Mixed finite element methods for the fully non- linear monge-ampère equation based on the vanishing moment method. SIAM Journal on Numerical Analysis, 2(47):1226–1250, 2009.

[30] M. Fortin and R. Glowinski. Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. Elsevier Science, 1983.

[31] B. D. Froese. A numerical method for the elliptic monge-ampère equation with transport boundary conditions. SIAM Journal on Scientific Computing, 34(3):A1432–A1459, 2012.

[32] H. W. Kuhn. The Hungarian method of solving the assignment problem. Naval Res. Logistics Quart., 2:83–97, 1955.

[33] 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(1):197–211, 2010.

[34] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. International Journal of Computer Vision, 60(3):225–240, 2004.

[35] F. H. Harlow and J. E. Welch. Numerical calculation of time-dependent viscous incompressible flow of fluid

with free surface. Physics of Fluids, 8(12):2182–2189, 1965.
[36] A. Iollo and D. Lombardi. A lagrangian scheme for the solution of the optimal mass transfer problem. Journal of Computational Physics, 230(9):3430–3442, 2011.
[37] Y-H. Kim and B. Pass. Multi-marginal optimal transport on riemannian manifolds. Technical report, Preprint arXiv:1303.6251, 2013.
[38] R. Kimmel and J.A. Sethian. Computing Geodesic Paths on Manifolds. Proc. of the National Academy of Sciences, 95(15):8431–8435, 1998.
[39] H. Ling and K. Okada. An efficient earth mover's distance algorithm for robust histogram comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(5):840–853, 2007.
[40] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. Siam J. Numer. Anal., 16:964–979, 1979.
[41] R.J. McCann. Polar factorization of maps on riemannian manifolds. Geometric & Functional Analysis, 11(3):589–608, 2001.
[42] Q. Mérigot. A multiscale approach to optimal transport. Computer Graphics Forum, 30(5):1583–1592, 2011.
[43] Y. E. Nesterov and A. S. Nemirovsky. Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms. SIAM Publishing, 1993.
[44] A. M. Oberman. Wide stencil finite difference schemes for the elliptic monge-ampere equation and functions of the eigenvalues of the hessian. Discrete and Continuous Dynamical Systems - Series B, 1(10):221–238, 2008.
[45] 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. Numerische Mathematik, 3(54):271–293, 1988.
[46] J. Rabin and G. Peyré. Wasserstein regularization of imaging problem. In IEEE International Conference on Image Processing (ICIP'11), pages 1541–1544, 2011.
[47] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In Scale Space and Variational Methods in Computer Vision (SSVM'11), volume 6667, pages 435–446, 2011.
[48] Y. Rubner, C. Tomasi, and L.J. Guibas. A metric for distributions with applications to image databases. In IEEE International Conference on Computer Vision (ICCV'98), pages 59–66, 1998.
[49] J. A. Sethian. A fast marching level set method for monotonically advancing fronts. Proc. of the National Academy of Sciences, 93(4):1591–1595, 1996.
[50] S. Shirdhonkar and D.W. Jacobs. Approximate earth mover's distance in linear time. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), pages 1–8, 2008.
[51] G. Strang. Linear Algebra and Its Applications. Brooks Cole, 1988.
[52] M. Sulman, J. F. Williams, and R. D. Russell. Optimal mass transport for higher dimensional adaptive grid generation. Journal of Computational Physics, 230(9):3302–3330, 2011.
[53] J. Tsitsiklis. Efficient Algorithms for Globally Optimal Trajectories. IEEE Trans. on Automatic Control, 40(9):1528–1538, Step. 1995.
[54] C. Villani. Topics in Optimal Transportation. Graduate Studies in Mathematics Series. American Mathematical Society, 2003.
[55] C. Villani. Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften. Springer, November 2008.