

Transport based image morphing with intensity modulation

Jan Maas¹, Martin Rumpf², and Stefan Simon²

¹ Institute of Science and Technology Austria, Klosterneuburg
jan.maas@ist.ac.at

² Institute for Numerical Simulation, Universität Bonn,
{martin.rumpf}@ins.uni-bonn.de, {s6stsim}@uni-bonn.de

Abstract We present a generalized optimal transport model in which the mass-preserving constraint for the L^2 -Wasserstein distance is relaxed by introducing a source term in the continuity equation. The source term is also incorporated in the path energy by means of its squared L^2 -norm in time of a functional with linear growth in space. This extension of the original transport model enables local density modulations, which is a desirable feature in applications such as image warping and blending. A key advantage of the use of a functional with linear growth in space is that it allows for singular sources and sinks, which can be supported on points or lines. On a technical level, the L^2 -norm in time ensures a disintegration of the source in time, which we use to obtain the well-posedness of the model and the existence of geodesic paths. The numerical discretization is based on the proximal splitting approach [18] and selected numerical test cases show the potential of the proposed approach. Furthermore, the approach is applied to the warping and blending of textures.

Keywords: Optimal transport, texture morphing, generalized Wasserstein distance, proximal splitting

1 An optimal transport model with source term

In the last decade optimal transport became a very popular tool in image processing and image analysis [18], where the quadratic Wasserstein distance is applied for instance in non-rigid image registration and warping. It was also used to robustly measure distances between images or to segment and classify images [17]. Driven by applications for instance in imaging [21,3,2] there is a strong demand to develop robust and efficient algorithms to compute optimal transport geodesics, such as the entropic regularization [4,16] or the sparse multiscale approach [22].

In their groundbreaking paper Benamou and Brenier [1] reformulated (for numerical purposes) the problem of optimal transport first considered by Monge and then relaxed by Kantorovich in a continuum mechanical framework describing the evolution of the mass distribution in time. This reformulation turned out to be the geodesic equation on the L^2 Wasserstein space. For an underlying flow of a density θ with Eulerian velocity v one considers the path energy

$$\mathcal{E}[\theta, v] = \int_0^1 \int_D \theta |v|^2 dx dt, \quad (1)$$

where $D \subset \mathbb{R}^d$ is assumed to be a closed, bounded convex domain with Lipschitz boundary. Then the quadratic Wasserstein distance $W_2[\theta_A, \theta_B]$ between two probability density function θ_A and θ_B can be computed by minimizing \mathcal{E} over all density functions $\theta : [0, 1] \times D \rightarrow \mathbb{R}$ and velocity fields $v : [0, 1] \times D \rightarrow \mathbb{R}^d$ subject to the continuity equation $\partial_t \theta + \operatorname{div}(\theta v) = 0$ and the constraints $\theta(0) = \theta_A$ and $\theta(1) = \theta_B$. Here the continuity equation enforces $\theta(t)$ to remain in the space of probability densities. In applications such as image registration or image morphing, input images are frequently not of the same mass. Thus, a contrast modulation on the input images is required before an optimal match between the

input images can be computed. But, even if the total mass of the input images coincides, the incorporation of local intensity modulation is desirable to cope with the variability of natural images and to avoid "artificial" long range transport just for the purpose of mass redistribution between totally independent image structures.

Recently, several optimal transport models [19,15,14,8,13] have been proposed, which relax the mass preserving condition and incorporate a source term in the transport model to allow arbitrary input measures. To this end one introduces a source term $z : [0, 1] \times D \rightarrow \mathbb{R}$ in the modified continuity equation

$$\partial_t \theta + \operatorname{div}(\theta v) = z. \quad (2)$$

This source terms has then to be incorporated in the path energy via a suitable penalty term, which represents the cost of mass production. Here, we propose the following generalized action functional

$$\mathcal{E}_\delta[\theta, v, z] = \int_0^1 \int_D \theta |v|^2 dx dt + \frac{1}{\delta} \int_0^1 \left(\int_D r(z) dx \right)^2 dt \quad (3)$$

subject to the relaxed continuity equation (2) and the constraints $\theta(0) = \theta_A$ and $\theta(1) = \theta_B$, where $r : \mathbb{R} \rightarrow \mathbb{R}$ is a non-negative, convex function satisfying $r(0) = 0$ and the linear growth condition $r(s) \leq C(1+|s|)$ for all $s \in \mathbb{R}$. Cases of interest in our considerations are $r(s) = |s|$ corresponding to the L^1 norm in space or more appropriate in imaging applications a Huber norm in space with $r(s) = \frac{1}{2\beta} s^2$ for $|s| \leq \beta$ and $|s| - \frac{\beta}{2}$ else for some $\beta > 0$. In this model $\delta > 0$ denotes a penalty parameter which allows to grade the mass modulation rate. It is desirable to allow also for singular sources which are supported on line segments of points in space. The linear growth property will ensure that singular source terms are admissible.

For this model existence of a shortest (geodesic) paths given by a minimizer of the cost functional \mathcal{E}_δ existed. To prove this the appropriate framework is that of Radon measures and a suitable decomposition of the measures for mass, momentum and source term into absolutely continuous and orthogonal parts with respect to the Lebesgue measure. Since these decompositions are not unique, it is useful to require 1-homogeneity of the integrands for the singular measures, which ensures that the definition of the energy functionals does not depend on the decomposition. In particular, we will observe that our class of models allows singular sources. Furthermore, the L^2 -norm in time provides an equi-integrability estimate, which guarantees compactness in the space of curves of Radon measures and is thus essential to establish existence of a minimizer.

The flow formulation (1) has been used in [1] primarily to compute optimal transport geodesics numerically with an augmented Lagrangian approach. In [18] it was shown that a proximal splitting algorithm leads to an equivalent optimization method. We will extend this approach to derive a suitable numerical discretization of our model.

This paper is organized as follows: First, in Section 2 we give an overview of recent developments on optimal transport models with source term. In Section 3 we rigorously define the generalized optimal transport model and establish the existence of optimal transport geodesics. Then we propose in Section 4 an efficient numerical scheme via proper adaptation of the proximal splitting method. Finally, in Section 5 we present results and discuss properties of the generalized model.

2 Related work on optimal transport with source term

In very recent years there has been a lot of activity on the extension of optimal transport distances to spaces of densities or measures with possibly different masses, which we here briefly summarize and point out differences to our model. A so-called partial optimal transport model was proposed by Caffarelli and McCann [5] and analyzed by Figalli [12]. By relaxing the marginal constraint in the Kantorovich

formulation, they ask for the optimal transport of a fixed fraction of some initial to a final density function. Note that there is actually no source term involved, but one is rather interested in the geometry of the subsets which are actually transported.

Furthermore, there are generalized transport distances which are closely related to the Benamou-Brenier formulation and based on a minimization of a path energy subject to the continuity equation with source term. Picolli and Rossi [19,20] considered minimizer of the path energy

$$\mathcal{E}_{L^1} = \int_0^1 \int_D \theta |v|^2 dx dt + \left(\int_0^1 \int_D |z| dx dt \right)^2$$

subject to equation (2). They prove for absolutely continuous measures θ and absolutely continuous sources z with respect to the Lebesgue measure that this geodesic formulation corresponds to solving the problem

$$\inf_{\tilde{\theta}_A, \tilde{\theta}_B: |\tilde{\theta}_A| = |\tilde{\theta}_B|} |\tilde{\theta}_A - \theta_A| + |\tilde{\theta}_B - \theta_B| + W_2(\tilde{\theta}_A, \tilde{\theta}_B).$$

Instead of the squared L^1 norm, Maas et al. [15] choose as penalization of the source term the squared L^2 norm:

$$\mathcal{E}_{L^2} = \int_0^1 \int_D \theta |v|^2 dx dt + \int_0^1 \int_D |z|^2 dx dt.$$

Chizat et al. [8] and Liero et al. [14] proposed an interpolating distance between the Wasserstein distance and the Fisher-Rao distance by minimizing the energy

$$\mathcal{E}_{WF} = \int_0^1 \int_D \theta (|v|^2 + \alpha(z)) dx dt$$

subject to the continuity equation $\partial_t \theta + \operatorname{div}(\theta v) = \theta z$. Note that the source term in this model is integrated w.r.t. to the measure given by θ . Furthermore, in [7] a static Kantorovich formulation is derived and it is shown that the distance in [19,20] arises as a special case.

At first glance, the distances obtained by minimizing the energies \mathcal{E}_{L^1} and \mathcal{E}_{L^2} seem to be very similar to our proposed energy (3). The difference becomes crucial when properly extending the energies to the space of Radon measures. In fact, a penalization of the source term squared L^2 norm does not allow singular sources which are for instance concentrated on lines or points, whereas such sources are possible in our model as it will be demonstrated in Section 4. However, choosing a penalization in the L^1 norm in space-time does not guarantee that the resulting minimizing measure is actually a curve in time in the space of Radon measures. Indeed, the generalized Benamou and Brenier model (3) allows for singular measures as source terms, which are for instance concentrated on lower dimensional sets.

3 Variational formulation for the generalized transport model

Here, we formulate a measure-valued setup for the energy in (3) as well as for the continuity equation with source term (2). We follow the lines of [9] and the treatment of the source term in the L^2 -norm in space-time presented in [15].

First, we apply the change of variables $(\theta, v) \mapsto (\theta, m = \theta v)$ already used by Benamou and Brenier [1]. Instead of the pair (θ, v) we consider the pair (θ, m) , where m denotes the momentum, such that the integrand $|v|^2 \theta$ pointwise transforms into

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

with the advantage that Φ is lower-semicontinuous, convex and 1-homogeneous.

A generalized continuity equation $\partial_t \mu + \operatorname{div}(\nu) = \zeta$ in terms of measure-valued quantities, namely mass $\mu \in \mathcal{M}^+([0, 1] \times D)$, momentum $\nu \in \mathcal{M}([0, 1] \times D; \mathbb{R}^d)$, and source term $\zeta \in \mathcal{M}([0, 1] \times D)$ and for given boundary values $\mu_0 = \mu_A$ and $\mu_1 = \mu_B$ is defined in the sense of distributions, by testing against all space-time test functions $\eta \in C^1([0, 1] \times D)$:

$$0 = \int_0^1 \left[\int_D \partial_t \eta(t, x) \, d\mu_t(x) + \int_D \nabla \eta(t, x) \cdot d\nu_t(x) + \int_D \eta(t, x) \, d\zeta_t(x) \right] dt - \int_D \eta(1, x) \, d\mu_B(x) + \int_D \eta(0, x) \, d\mu_A(x).$$

We will consider curves of measures on D instead of just measures on the product space $[0, 1] \times D$ as the proper measure theoretic setup for the continuity equation with source term. Further, we denote the set of all solutions of the weak continuity equation with source term by $\mathcal{CE}[0, 1]$.

Next, we define the energy (3) in terms of measures. To this end we decompose for each $t \in [0, 1]$ the triple $(\mu_t, \nu_t, \zeta_t) \in \mathcal{M}^+(D) \times \mathcal{M}(D; \mathbb{R}^d) \times \mathcal{M}(D)$ using the Lebesgue decomposition theorem. Thus, we can rigorously define the total energy functional \mathcal{E}_δ (cf. 3) for measures $(\mu, \nu, \zeta) \in \mathcal{M}^+([0, 1] \times D) \times \mathcal{M}([0, 1] \times D; \mathbb{R}^d) \times \mathcal{M}([0, 1] \times D)$ using these decompositions. Finally, following [15], we obtain the following existence result for minimizing paths.

Theorem 1 (Existence of geodesics). *Let $\delta \in (0, \infty)$ and take $\mu_A, \mu_B \in \mathcal{M}^+(D)$. Then there exists a minimizer $(\bar{\mu}_t, \bar{\nu}_t, \bar{\zeta}_t)_{t \in [0, 1]}$ of the energy \mathcal{E}_δ subject to the weak continuity equation (4). Moreover, this defines a metric \mathcal{W}_δ on $\mathcal{M}^+(D)$, and the associated curve $(\bar{\mu}_t)_{t \in [0, 1]}$ is a constant speed geodesic for \mathcal{W}_δ , i.e.,*

$$\mathcal{W}_\delta[\bar{\mu}_s, \bar{\mu}_t] = |s - t| \mathcal{W}_\delta[\mu_A, \mu_B]$$

for all $s, t \in [0, 1]$.

The proof of this theorem with an additional representation formula for the generalized Wasserstein distance is given in the supplementary material accompanying this paper.

4 Proximal splitting algorithm

In this section we derive a numerical scheme to compute geodesics for our new distance for $d = 2$. To this end, we will adapt the proximal splitting algorithm, which was proposed by Papadakis et al. [18] for the classical L^2 optimal transport problem. In detail, the constraint optimization problem is first rewritten as a non-constraint minimization problem adding the indicator function of the set of solutions of the continuity equation $\mathcal{CE}[0, 1]$ to the cost functional. Then, the proximal splitting algorithm yields a solution scheme, which only requires to solve a space-time elliptic problem and to project pointwise onto a convex set. The resulting algorithm is equivalent to the augmented Lagrangian approach in [1]. Different from [1, 18] we will use a finite element discretization instead of finite differences.

Let us briefly recall the definition and the basic properties of a proximal mapping (see for instance [6, 18]). In the following, let $(X, \|\cdot\|_X)$ be a Hilbert space and $f : X \rightarrow \mathbb{R} \cup \{\infty\}$ a convex and lsc function. Then the proximal mapping of f is defined as

$$\operatorname{prox}_f(x) = \operatorname{argmin}_{y \in X} f(y) + \frac{1}{2} \|x - y\|^2.$$

In the sequel it will be important to compute the proximal mapping of the indicator function

$$\mathcal{I}_K(x) = \begin{cases} 0 & \text{if } x \in K, \\ +\infty & \text{if } x \notin K. \end{cases}$$

of a convex set $K \subset X$, which is just given by $\text{prox}_{\mathcal{I}_K}(x) = \text{proj}_K(x)$, where proj_K is the orthogonal projection on K with respect to the norm $\|\cdot\|_X$. Now, we suppose that D is a polygonal domain and consider a tetrahedral mesh S_h with grid size h for the space time domain $[0, 1] \times D$, which is generated from a triangular mesh for the domain D via subdivision of prisms $(k h, (k+1)h) \times T$ (with T being a triangle) into 3 tetrahedrons such that the resulting tetrahedral mesh is an admissible triangulation in space time. On this triangulation we define finite element spaces

$$\begin{aligned} V^1(S_h) &= \{\phi_h : [0, 1] \times D \rightarrow \mathbb{R} : \phi_h \text{ continuous and piecewise linear on elements in } S_h\}, \\ V^0(S_h) &= \{\theta_h : [0, 1] \times D \rightarrow \mathbb{R} : \theta_h \text{ piecewise constant on elements in } S_h\}. \end{aligned}$$

This allows us define discretization

$$\theta_h \in V^0(S_h), \quad m_h \in (V^0(S_h))^d, \quad z_h \in V^1(S_h),$$

for the measures for mass, momentum and source, respectively. Furthermore, we will use the notation $p_h = (\theta_h, m_h) \in V^0(S_h)^{d+1}$. For a triple $(\theta_h, m_h, z_h) \in V_h^0(S) \times V_h^0(S)^d \times V_h^1(S)$ we choose a weighted L^2 norm

$$\|(\theta_h, m_h, z_h)\| := \left(\int_0^1 \int_D |\theta_h|^2 + |m_h|^2 + \frac{1}{\delta} |z_h|^2 dx dt \right)^{\frac{1}{2}},$$

which can be computed exactly by choosing a quadrature rule of at least second order. In correspondence to the weak formulation (4) the set of discrete solutions of a continuity equation is defined as follows:

Definition 1. Let $\theta_A, \theta_B \in V^0(S_h)$ be given. Then, the set \mathcal{CE}_h of solutions of a weak continuity equation with source term and boundary values θ_A, θ_B is given by all triples $(\theta_h, m_h, z_h) \in V_h^0(S) \times V_h^0(S)^d \times V_h^1(S)$ satisfying

$$\int_0^1 \int_D \theta_h \partial_t \phi_h + m_h \nabla_x \phi_h + z \phi_h dx dt = \int_D \phi_h(1) \theta_B - \phi_h(0) \theta_A dx \quad \forall \phi_h \in V^1(S_h).$$

Note that we used Neumann boundary condition in space. The approach can easily be adopted in case of Dirichlet or periodic boundary conditions.

Now, we can state a discrete version of the minimization problem:

$$\inf_{(\theta_h, m_h, z_h) \in \mathcal{CE}_h} \left(\int_0^1 \int_D \Phi(\theta_h, m_h) dx dt + \frac{1}{\delta} \int_0^1 \left(\int_D \mathcal{R}_h(z_h) dx \right)^2 dt \right),$$

where $\mathcal{R}_h[z_h]$ denotes a suitable interpolation of $r(z_h)$. Here, we define $\mathcal{R}_h[z_h](t, x)$ as the piecewise affine interpolation of $r(z_h((k-1)h, \cdot))$ on the triangle T for $(t, x) \in (k h, (k+1)h) \times T$ (one of the prisms underlying the tetrahedral grid). Numerically, we are not able to treat singular measures as presented in Section 3. Our concrete choices of $r(s)$, which coincide with $|s|$ for large s allow to approximate such measures in the source term cost supported on the union of the support of basis functions. Thus point or line sources are numerically treated via sources with support thickness $2h$. To apply a proximal splitting

algorithm, we split the functional into

$$F_1(\theta_h, m_h, z_h) := F_{\text{trans}}(\theta_h, m_h) + \frac{1}{\delta} F_{\text{source}}(z_h)$$

$$\text{with } F_{\text{trans}}(\theta_h, m_h) := \int_0^1 \int_D \Phi(\theta_h, m_h) \, dx \, dt, \quad F_{\text{source}}(z_h) := \int_0^1 \left(\int_D \mathcal{R}_h[z_h] \, dx \right)^2 \, dt,$$

$$F_2(\theta_h, m_h, z_h) := \mathcal{I}_{\mathcal{CE}_h}(\theta_h, m_h, z_h).$$

Next, let us compute the proximal mappings of F_1 and F_2 .

Proximal map of F_2 . The computation of the proximal mapping of the indicator function of \mathcal{CE}_h requires the orthogonal projection of a point $(p_h = (\theta_h, m_h), z_h) \in V^0(S)^{d+1} \times V^1(S)$ onto \mathcal{CE}_h , i.e. we ask for $(p_h^*, z_h^*) \in \operatorname{argmin}_{(q_h, w_h) \in \mathcal{CE}_h} \|(p_h, z_h) - (q_h, w_h)\|^2$. The associated Lagrangian is given by

$$L[(q_h, w_h), \psi_h] = \|(p_h, z_h) - (q_h, w_h)\|^2 - \int_0^1 \int_D q_h \cdot \nabla_{(t,x)} \psi_h + w_h \psi_h \, dx \, dt + \int_D \psi_h(1) \theta_B - \psi_h(0) \theta_A \, dx,$$

with a Lagrange multiplier $\psi_h \in V^1(S_h)$. In terms of this Lagrangian the projection problem can be written as a saddle point problem, where as ask for $(p_h^*, z_h^*, \phi_h^*) \in V^0(S)^{d+1} \times V^1(S) \times V^1(S)$, such that

$$L[(p_h^*, z_h^*, \phi_h^*)] = \min_{(q_h, w_h) \in V_h^0(S)^{d+1} \times V_h^1(S)} \max_{\psi_h \in V_h^1(S)} L[(q_h, w_h), \psi_h].$$

The Euler-Lagrange equations corresponding to this saddle point problem are given by

$$\int_0^1 \int_D p_h^* \cdot \nabla_{(t,x)} \psi_h + z_h^* \psi_h \, dx \, dt = \int_D \psi_h(1) \theta_B - \psi_h(0) \theta_A \, dx \quad \forall \psi_h \in V_h^1(S) \quad (6)$$

$$\int_0^1 \int_D q_h \cdot \nabla_{(t,x)} \phi_h^* \, dx \, dt = \int_0^1 \int_D 2(p_h^* - p_h) q_h \, dx \, dt \quad \forall q_h \in V_h^0(S)^{d+1} \quad (7)$$

$$\int_0^1 \int_D \phi_h^* w_h \, dx \, dt = \int_0^1 \int_D \frac{2}{\delta} (z_h^* - z_h) w_h \, dx \, dt \quad \forall w_h \in V_h^1(S) \quad (8)$$

Testing equation (7) with $q_h = \nabla_{(t,x)} \psi_h$ and then using equation (6) gives

$$\begin{aligned} \int_0^1 \int_D \frac{1}{2} \nabla_{(t,x)} \phi_h^* \cdot \nabla_{(t,x)} \psi_h \, dx \, dt &= \int_0^1 \int_D (p_h^* - p_h) \cdot \nabla_{(t,x)} \psi_h \, dx \, dt \\ &= \int_D \psi_h(1) \theta_B - \psi_h(0) \theta_A \, dx - \int_0^1 \int_D z_h^* \psi_h + p_h \cdot \nabla_{(t,x)} \psi_h \, dx \, dt \end{aligned}$$

Hence, by using equation (8) ($z_h^* = z_h + \frac{\delta}{2} \phi_h^*$) we obtain

$$\begin{aligned} \int_0^1 \int_D \frac{1}{2} \nabla_{(t,x)} \phi_h^* \cdot \nabla_{(t,x)} \psi_h + \frac{\delta}{2} \phi_h^* \psi_h \, dx \, dt &= \int_D \psi_h(1) \theta_B - \psi_h(0) \theta_A \, dx \\ &\quad - \int_0^1 \int_D z_h \psi_h + p_h \cdot \nabla_{(t,x)} \psi_h \, dx \, dt \end{aligned}$$

for all $\psi_h \in V_h^1(S)$.

After computing ϕ_h the solution of the projection problem is given by

$$p_h^* = p_h + \frac{1}{2} \nabla_{(t,x)} \phi_h^*, \quad z_h^* = z_h + \frac{\delta}{2} \phi_h^*.$$

Proximal map of F_1 . The transport term F_{trans} does only depend on θ_h and m_h and can be treated exactly as for classical optimal transport. Since we observe pointwise that $\Phi^* = \mathcal{I}_K$ is an indicator function of the convex set

$$K = \left\{ (\theta, m) : \theta + \frac{|m|^2}{4} \leq 0 \right\},$$

(see [1]) we can use Moreau's identity and get

$$\text{prox}_{\Phi}(\theta, m) = (\theta, m) - \text{prox}_{\Phi^*}(\theta, m) = (\theta, m) - \text{proj}_K(\theta, m).$$

The projection onto K can be computed separately on each tetrahedron of the simplicial mesh S_h due to the choice of our finite element spaces with $p_h \in V^0(S_h)^{d+1}$.

We note that for a source term in L^2 both in space and time we easily get a pointwise update

$$\text{prox}_{\frac{\gamma}{2}|\cdot|^2}(z) = \underset{w}{\text{argmin}} \frac{1}{\delta} |w|^2 + \frac{1}{\delta} |w - z|^2 = \frac{1}{1 + \gamma} z.$$

Following the computation in [11] we also get a pointwise update for the proximal operator of the source term in $L^1(L^1)$, which is given by

$$\text{prox}_{\frac{\gamma}{2}|z|}(z) = \begin{cases} 0, & \text{if } |z| \leq \frac{\gamma}{2} \\ z - \frac{\gamma}{2} \text{sgn}(z), & \text{else.} \end{cases}$$

Thus, a numerical scheme for a source term in $L^1(L^1)$ would be as simple as for a source term in $L^2(L^2)$, but existence of geodesics is not guaranteed. In case of a linear growth function $r(\cdot)$ the minimization problem only decouples in time but not in space. Hence, for each discrete time step k we have to solve

$$\underset{w_h}{\text{argmin}} \frac{\gamma}{\delta} \left(\int_D \mathcal{R}_h[w_h](kh, x) dx \right)^2 + \frac{1}{2\delta} \int_D |w_h(kh, x) - z_h(kh, x)|^2 dx.$$

For a source term in $L^2(L^1)$ the minimization problem is well defined, but since $r(z) = |z|$ is not differentiable it is not clear how to find the minimizer. Therefore we restrict our numerical computations to the case of r being the Huber function and use a gradient descent to compute this minimum.

Douglas-Rachford splitting algorithm. Finally, to solve the minimization problem

$$(p_h^*, z_h^*) \in \underset{(q_h, w_h) \in V^0(S)^{d+1} \times V^1(S)}{\text{argmin}} F_1(q_h, w_h) + F_2(q_h, w_h)$$

we apply the Douglas-Rachford splitting algorithm [10], which is given by the iteration

$$\begin{aligned} (q_h^n, w_h^n) &= \text{prox}_{\gamma F_2}((p_h^{n-1}, z_h^{n-1})), \\ (p_h^n, z_h^n) &= (p_h, z_h)^{n-1} + \alpha \left(\text{prox}_{\gamma F_1}(2(q_h^n, w_h^n) - (p_h, z_h)^{n-1}) - (q_h, w_h)^n \right), \end{aligned}$$

for an initial value (p_h^0, z_h^0) and a step size weight α have to be chosen. It is guaranteed that for $\gamma > 0$ and $\alpha \in (0, 2)$ the sequence (p_h^n, z_h^n) as well as $(q_h^n, w_h^n)^n$ converges to a solution of the minimization problem.

5 Numerical results

We have applied the proposed scheme for the optimal transport with source term for different sets of (θ_A, θ_B) , which we show in the accompanying figures always from left ($\theta(0) = \theta_A$) to right ($\theta(1) = \theta_B$). In all computations $D = (0, 1)^2$ and the grid size is $h = 2^{-7}$. Recall, that the Huber function is given by $r(s) = \frac{1}{2\beta} s^2$ for $|s| \leq \beta$ and $|s| - \frac{\beta}{2}$, where we choose $\beta = 10^{-4}$. At first, we demonstrate that the density function r for the source term is the right choice to deal with approximations of singular measures as source terms. To this end, we consider in Figure 1 measures θ_A and θ_B supported on a thin rectangular strip with constant but different density. The proposed model with the L^2 -Huber [$L^2(H)$] type cost functional $\int_0^1 (\int_D r(\theta) dx)^2 dt$ for the source term is able to generate the required singular measure and the Wasserstein geodesic is just given by a blending of the two measure θ_A and θ_B . The generating of singular sources is not possible for an L^2 type cost functional in space time [$L^2(L^2)$], which was proposed in [15]. Indeed, chosen the cost functional $\int_0^1 \int_D z^2 dx dt$ for the same data, the generation of mass via the source term takes place on a thick super set of the rectangular strip and is then transported toward to strip. We also observe a similar effect for absolutely continuous measures. In Figure 2 we compare the $L^2(H)$ and the $L^2(L^2)$ source term for geodesics connecting differently scaled characteristic functions of a square. Again, the resulting geodesic for the $L^2(H)$ -model is given by a blending of the two measure θ_A and θ_B , whereas in the $L^2(L^2)$ the additional mass in image u_B is generated from a bigger support. Figure 3 shows a plot of $t \mapsto \int_D |z(t, \cdot)| dx$ for both models underlining the equidistribution of the source in time for the $L^2(H)$ model. Let us remark, that numerical diffusion in particular on coarse meshes leads to a blurring effect for the source term at discontinuities of the density which is then accompanied by minor transport to compensate for this diffusion.



Figure 1. Optimal transport geodesic between approximations of singular measures with different intensity. Here the source term parameter is $\delta = 10^0$.

Next, we investigate the effect of the source term parameter δ for the $L^2(H)$ model. In Figure 4 we choose as input data θ_A at time $t = 0$ a characteristic function of a square and as input data θ_B at time $t = 1$ the same measure density with an additional characteristic function of a translated square of the same size. Now, optimizing the connecting path with respect to the generalized Wasserstein distance there is a competition between the curve which simply blends the second square and a curve which transports part of the second square and blends of remaining non transported measure. This balance between both processes depends on δ . In the limit $\delta \rightarrow 0$ transport becomes cheaper, which is reflected by the computational results for small δ . In contrast for $\delta \rightarrow \infty$ transport becomes expensive and a simple blending can be observed for large values of δ in Figure 4. A similar effect is shown in Figure 5, where the a bump function in the center of a periodic cell is transported to a splitted bump function in the corners applying periodic boundary conditions.

In Figure 6 another type of interaction between generation and transport of mass is shown. The initial images at time $t = 0$ consists of three scaled characteristic functions of balls, where one of this balls has

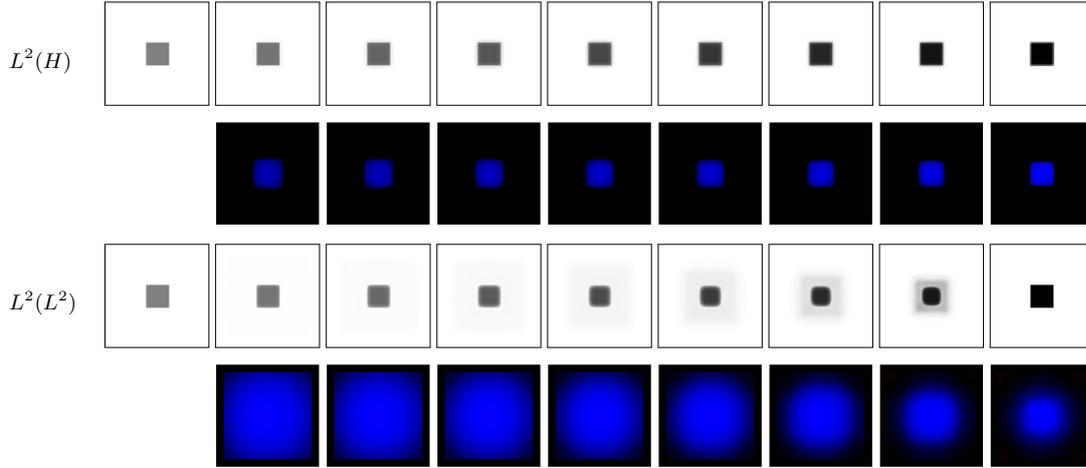


Figure 2. Optimal transport geodesic and corresponding source terms between two characteristic functions of squares with different intensity. Here the source term parameter is $\delta = 10^0$.

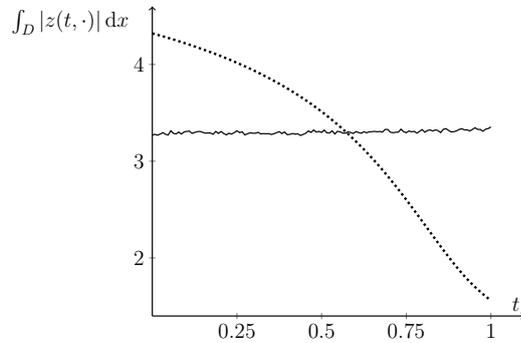


Figure 3. Distribution of the L^1 -norm of the source term in time for the example in Figure 2. (continuous line: $L^2(H)$, dotted line: $L^2(L^2)$)

smaller density value. The final image at time $t = 1$ is based on the identical geometric configuration, but with swapped densities. Depending on the parameter δ a certain amount of mass is transported from the two balls with higher intensity in the image at time $t = 0$. At the same time a blending of the transported masses as a compensation for the non balanced total mass can be observed. Figure 7 shows plots of the functions $t \mapsto \int_D |z(t, \cdot)| dx$, $t \mapsto \int_D z^+(t, \cdot) dx$, and $t \mapsto \int_D |z^-(t, \cdot)| dx$ for the different values of δ .

A striking observation in Figure 3 and Figure 7 ist that $t \mapsto \int_D |z(t, \cdot)| dx$ is approximately constant in time for the $L^2(H)$ model. This is in contrast to the $L^2(L^2)$ model as indicated in Figure 3.

Finally Figure 8 and Figure 9 depict examples for images of wood texture and marble texture. Generalized Wasserstein geodesics in case of the $L^2(H)$ cost functional are shown.

References

1. Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.
2. Martin Burger, Marzena Franek, and Carola-Bibiane Schönlieb. Regularized regression and density estimation based on optimal transport. *Applied Mathematics Research eXpress*, 2012(2):209–253, 2012.

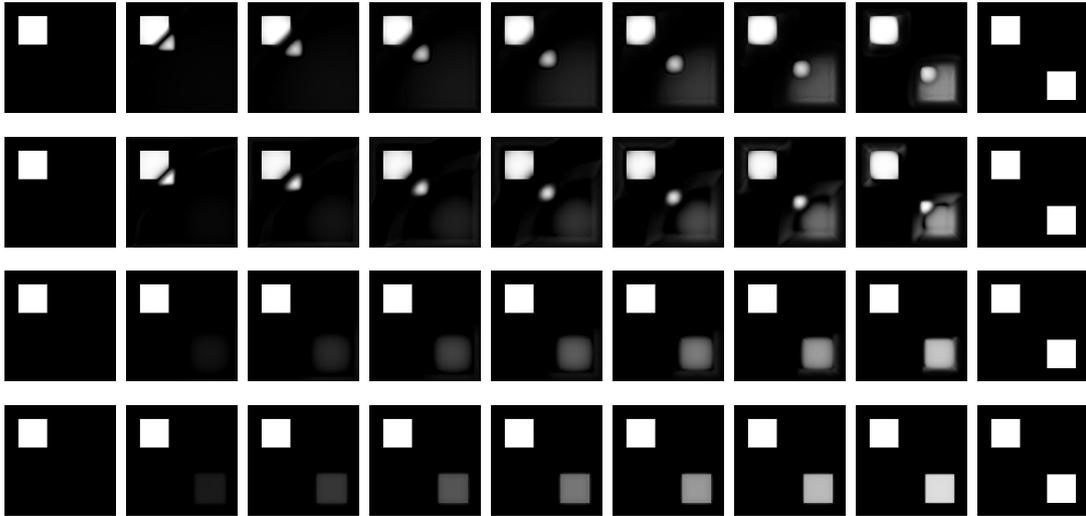


Figure 4. Generalized Wasserstein geodesic between a scaled characteristic functions on a square and two differently scaled characteristic functions of squares of the same size. The dependence on the source term parameter is shown for the $L^2(H)$ model. From top to bottom $\delta = 10^{-2}, 10^{-1}, 10^0, 10^1$.

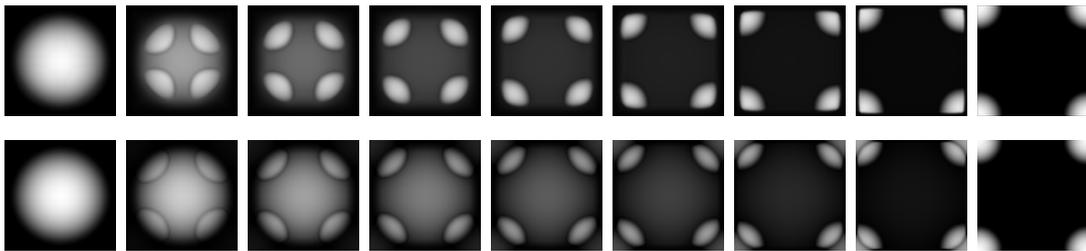


Figure 5. Generalized Wasserstein geodesic connecting two translated bump functions are computed. The bump functions are periodically extended and $\delta = 10^{-2}$ (top) and $\delta = 10^0$ (bottom).

3. Giuseppe Buttazzo and Filippo Santambrogio. A model for the optimal planning of an urban area. *SIAM J. Math. Anal.*, 37(2):514–530, 2005.
4. Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *arXiv preprint arXiv:1512.02783*, 2015.
5. Luis A Caffarelli and Robert J McCann. Free boundaries in optimal transport and monge-ampere obstacle problems. *Annals of mathematics*, pages 673–730, 2010.
6. Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. pages 185–212, 2011.
7. Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: geometry and kantorovich formulation. *arXiv preprint arXiv:1508.05216*, 2015.
8. Lenaic Chizat, Bernhard Schmitzer, Gabriel Peyré, and François-Xavier Vialard. An interpolating distance between optimal transport and fischer-rao. *arXiv preprint arXiv:1506.06430*, 2015.
9. Jean Dolbeault, Bruno Nazaret, and Giuseppe Savaré. A new class of transport distances between measures. *Calc. Var. Partial Differential Equations*, 34(2):193–231, 2009.
10. Jonathan Eckstein and Dimitri P Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

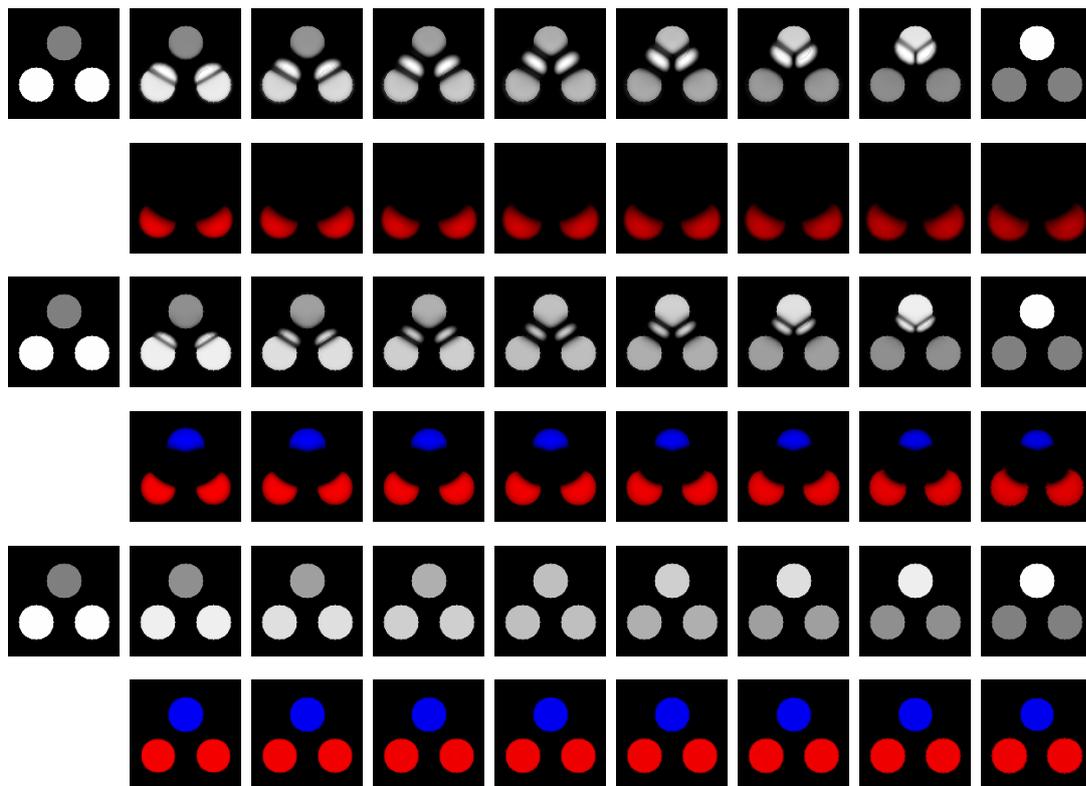


Figure 6. Optimal transport geodesic with corresponding distribution of the L^1 -norm of the source term in time between three scaled characteristic functions of balls with different densities. Here the source term parameter are from top to bottom $\delta = 10^0, 10^1, 10^2$.

11. Ernie Esser. Applications of lagrangian-based alternating direction methods and connections to split bregman. *CAM report*, 9:31, 2009.
12. Alessio Figalli. The optimal partial transport problem. *Archive for rational mechanics and analysis*, 195(2):533–560, 2010.
13. Stanislav Kondratyev, Léonard Monsaingeon, and Dmitry Vorotnikov. A new optimal transport distance on the space of finite radon measures. *arXiv preprint arXiv:1505.07746*, 2015.
14. Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal transport in competition with reaction: The Hellinger-Kantorovich distance and geodesic curves. Preprint no 2160, WIAS, 2015.
15. Jan Maas, Martin Rumpf, Carola Schönlieb, and Stefan Simon. A generalized model for optimal transport of images including dissipation and density modulation. *ESAIM Math. Model. Numer. Anal.*, 49(6):1745–1769, 2015.
16. Gabriel Peyré. Entropic wasserstein gradient flows. *arXiv preprint arXiv:1502.06216*, 2015.
17. Gabriel Peyré, Jalal Fadili, and Julien Rabin. Wasserstein active contours. In *Proc. of IEEE International Conference on Image Processing*, pages 2541–2544, 2012.
18. Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.
19. Benedetto Piccoli and Francesco Rossi. On properties of the generalized wasserstein distance. *arXiv preprint arXiv:1304.7014*, 2013.
20. Benedetto Piccoli and Francesco Rossi. Generalized wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211:335–358, 2014.

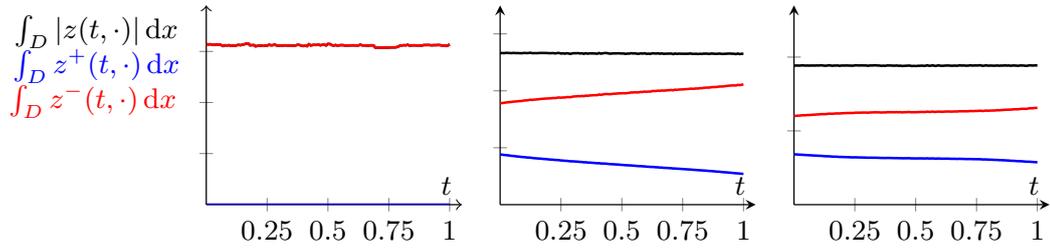


Figure 7. Distribution of the L^1 -norm of the source term in time for the example in Figure 6 for $\delta = 10^0, 10^1, 10^2$.

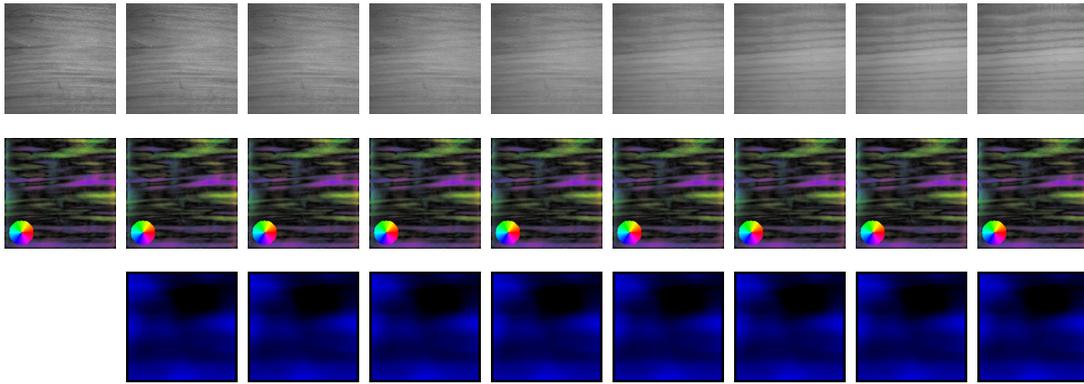


Figure 8. Optimal transport geodesic between wood textures (top). Here the source term parameter is $\delta = 10^0$. Further the corresponding momentum (middle) and source term (bottom) are depicted.

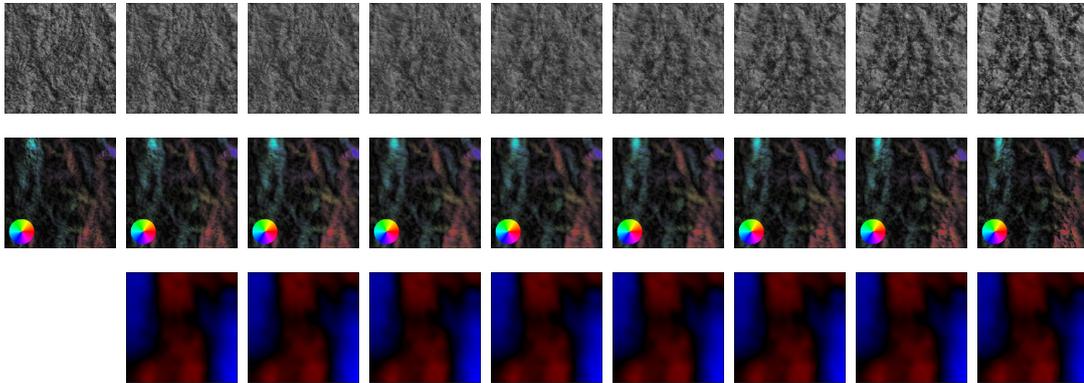


Figure 9. Optimal transport geodesic between marble textures (top). Here the source term parameter is $\delta = 10^{-2}$. Further the corresponding momentum (middle) and source term (bottom) are depicted.

21. Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.
22. Bernhard Schmitzer. A sparse multi-scale algorithm for dense optimal transport. *arXiv preprint arXiv:1510.05466*, 2015.