A generalized model for optimal transport of images including dissipation and density modulation

Jan Maas, Martin Rumpf, Carola Schönlieb, Stefan Simon

April 9, 2015

Abstract

In this paper the optimal transport and the metamorphosis perspectives are combined. For a pair of given input images geodesic paths in the space of images are defined as minimizers of a resulting path energy. To this end, the underlying Riemannian metric measures the rate of transport cost and the rate of viscous dissipation. Furthermore, the model is capable to deal with strongly varying image contrast and explicitly allows for sources and sinks in the transport equations which are incorporated in the metric related to the metamorphosis approach by Trouvé and Younes. In the non-viscous case with source term existence of geodesic paths is proven in the space of measures. The proposed model is explored on the range from merely optimal transport to strongly dissipative dynamics. For this model a robust and effective variational time discretization of geodesic paths is proposed. This requires to minimize a discrete path energy consisting of a sum of consecutive image matching functionals. These functionals are defined on corresponding pairs of intensity functions and on associated pairwise matching deformations. Existence of time discrete geodesics is demonstrated. Furthermore, a finite element implementation is proposed and applied to instructive test cases and to real images. In the non-viscous case this is compared to the algorithm proposed by Benamou and Brenier including a discretization of the source term. Finally, the model is generalized to define discrete weighted barycentres with applications to textures and objects.

Introduction

In the past two decades concepts from finite dimensional classical geometry have been successfully transferred to infinite-dimensional spaces, where shapes are contour curves of geometric objects, surfaces, image intensity maps, or probability densities. These concepts have a continuously increasing impact on the development of novel computational tools in computer vision and imaging, ranging from shape morphing and modeling [KMP07], and shape statistics, *e.g.* [FLPJ04], to texture analysis [RPDB12] and computational anatomy [BMTY02]. Three particularly influential approaches on the space of image maps are linked to optimal transportation [Mon81, BB00, ZYHT07, PPO14], the flow of diffeomorphism [Arn66, DGM98a] and metamorphosis [MY01, TY05]. Here, we combine these three approaches and explore properties of the resulting image manifold. Thus, in what follows we briefly review the underlying concepts.

Optimal transport and its application in imaging. The problem of optimal transport is introduced in the seminal work of Monge [Mon81] in 1781. In [Kan42, Kan48] Kantorovich proposes a relaxed formulation of Monge's problem which gives rise to the Wasserstein distance considered in this paper. Let (D, d) constitute a metric space. The 2-Wasserstein distance between two probability measures $\mu_A, \mu_B \in \mathbb{P}(D)$ is defined by

$$\mathcal{W}(\mu_A, \mu_B)^2 := \min_{\pi \in \Gamma(\mu_A, \mu_B)} \int_{D \times D} d(x, y)^2 \, \mathrm{d}\pi(x, y). \tag{1}$$

Here $\Gamma(\mu_A, \mu_B)$ denotes the set of all probability measures on $D \times D$ with marginals μ_A and μ_B with respect to x and y, respectively. For an introduction to optimal transport and the Wasserstein distance we refer the reader to the reviews [Amb03, Eva99, Vil03, AGS06, Vil08].

Being a distance function applicable to very general measures (continuous and discrete measures) the Wasserstein distance has an increasing impact on robust distance measures in imaging [RTG00, BS05, PFR12, BFS12]. As the Wasserstein metric is defined for arbitrary probability measures with finite second moment it allows to measure distances between absolutely continuous measures with respect to the Lebesgue measure as well as concentrated measures. With the increase of the complexity of applications efficient numerical computation of (3) became increasingly important. In that respect Benamou and Brenier propose an alternative formulation of the quadratic Wasserstein distance using the perspective of the underlying flow of a density θ with Eulerian velocity v and expressing the transport in terms of a

constrained flow [BB00]. That is, one asks for a minimizer of the path energy

$$\mathcal{E}[\theta, v] = \int_0^1 \int_D \theta |v|^2 \, \mathrm{d}t \tag{2}$$

for a density function $\theta : [0,1] \times D \to \mathbb{R}$ and a velocity field $v : [0,1] \times D \to \mathbb{R}^d$ subject to the transport equation $\partial_t \theta + \mathrm{div}(\theta v) = 0$ and the constraints $\theta(0) = \theta_A$ and $\theta(1) = \theta_B$. Then the minimal energy is indeed the squared Wasserstein distance between measures μ_A and μ_B with corresponding densities θ_A and θ_B respectively. Their algorithm has been immensely influential in the numerical computation of the Wasserstein distance and gradient flows related to it, see *e.g.* [BCW10, DMM10, PPO14]. In [PPO14], for instance, a proximal point algorithm for the solution of Benamou-Brenier's formulation (2) is derived and applied for the computation of Wasserstein geodesics between two image densities. Alternatively, if $D \subseteq \mathbb{R}^d$ is a strictly convex domain, d is the Euclidean distance, and μ_A and μ_B are absolutely continuous measures with densities θ_A and θ_B, respectively, then one has

$$\mathcal{W}(\mu_A, \mu_B)^2 = \min_{\phi_\# \mu_A = \mu_B} \int_D |\phi(x) - x|^2 \theta_A(x) \, \mathrm{d}x, \tag{3}$$

where $\phi_\# \mu_A$ denotes the push forward of the measure μ_A under the mapping ϕ and for a diffeomorphism ϕ the constraint can be expressed as $(\det D\phi) \, \theta_B \circ \phi = \theta_A$. In [AHT03] an initial mass preserving transport map [Mos65] is created and then an explicit time stepping scheme is employed to compute the optimal map from a modified formulation of (3) where the constraint is linearized. In [HRT10] the authors pick up formulation (3) as well and use a sequential quadratic programming method for its optimisation. Moreover, in [CWVB09] the authors propose a gradient descent for the dual formulation of (3) and show its use for image registration and warping. In [LR05, SAK10] a damped Newton method is used to compute a solution Ψ of the Monge-Ampere equation (which is the equality constraint on $\phi = \nabla \Psi$) and subsequently the optimal transport map ϕ. Finally, let us mention that in [SS13a] the authors propose another interesting numerical algorithm for the efficient computation of the Wasserstein distance that is based on an extension of the auction algorithm. The latter optimizes the cost functional in the Wasserstein distance only on a sparse subset of possible assignment pairs still guaranteeing global optimality. Various alternative computational approaches for the Wasserstein distance exist, *e.g.* [DG06, Obe08, PPO14].

In terms of imaging application the Wasserstein distance has been employed in the context of image and shape classification, segmentation, registration and warping, and image smoothing. In [RTG00] the Wasserstein distance is used as a distance on the images directly, interpreting images as discrete measures. In other approaches the Wasserstein distance is used on image histograms or points clouds (which could be feature vectors of images), see e.g. [GD04, LO07]. In the context of image segmentation and classification similar approaches are used, see, *e.g.* [CEN07, NBCE09, PFR12, OJBS12]. In [RPDB12] a discrete Wasserstein distance is employed for the computation of barycentres of discrete probability distributions with applications to texture synthesis and mixing. Thereby the original Wasserstein metric is replaced by a sliced version over one-dimensional distributions. In [RPC10] the same approach is used on clouds of geodesic shape descriptors as a similarity measure to discriminate between different shapes in 2D and 3D shape retrieval. The Wasserstein distance is also used in the context of contrast and colour modification in, *e.g.* [RP11, FPR+13]. In [ZHT03, HZTA04] the quadratic Wasserstein distance is considered to define a rigorous distance between images, applied to non-rigid image registration and warping. As a distance function for shapes the Gromov-Wasserstein distance is introduced in a series of works by Memoli [Mem07, Mém11]. Here, shapes are modelled as compact metric spaces and the Gromov-Wasserstein distance is computed on isometry classes of each space. The Gromov-Wasserstein differs from the Wasserstein distance (1) as it assigns a cost to pairs of transport assignments. In the context of surface dissimilarity measurement the Wasserstein distance is applied to metric densities on the hyperbolic disc representing conformal mappings of different surfaces [LD11]. In [SS13b] the authors create a convex shape-prior from a modified Gromov-Wasserstein distance. Their approach can be used for image segmentation problems in which prior shape knowledge on the objects that should be segmented can be provided in terms of a template shape. This approach is modified in [SS13c] using the quadratic Wasserstein distance as a regularizer between learned reference shapes and the segmentation. In [BFS12] the Wasserstein distance is used as a data fidelity term in a generic regularization approach and applied for image density estimation and cartoon-texture decomposition. Moreover, a review of the use of geodesic methods and in particular optimal transportation in computer vision can be found in [PPKC10].

Flow of diffeomorphism. The physical modeling of viscous flow involves dissipation as an integrated measure of local friction. Arnold [Arn66, AK98] proposes to study viscous flows from the perspective of a family of diffeomorphisms $(\phi(t))_{t \in [0,1]} : \bar{D} \to D$ which describe the transport of densities, *e.g.* image intensities, along particle paths $(\phi(t,x))_{t \in [0,1]}$ for $x \in D$. This concept is picked up in vision by Grenander and coworkers [Gre81, DGM98b]. As a Riemannian metric $\mathcal{G}^{\mathrm{diff}}[\cdot, \cdot]$ one considers the rate of viscous dissipation induced by the Eulerian flow velocity $v(t) = \dot{\phi}(t) \circ \phi^{-1}(t)$ in a

multipolar fluid model (cf. Nečas and Šilhavý [Nv91]). Here, the Eulerian motion field v is considered as a tangent vector on the manifold of diffeomorphisms. The resulting Riemannian metric one obtains $\tilde{\mathcal{G}}^{\mathrm{diff}}[v,v] := \int_D L[v,v]\,\mathrm{d}x$, where $L[v,v] = C\varepsilon[v] : \varepsilon[v]$ with $\varepsilon[v] = (\nabla v + (\nabla v)^T)/2$, and a deduced path energy $\tilde{\mathcal{E}}^{\mathrm{diff}}[\phi] = \int_0^1 \tilde{\mathcal{G}}^{\mathrm{diff}}[v,v]\,\mathrm{d}t$ as an action functional on flows encodes the total accumulated dissipation on the domain D and on the time interval $[0,1]$. To study the warping of two image intensity functions $\theta_A, \theta_B : D \to \mathbb{R}$ for which there exists a diffeomorphism ϕ with $\theta_B = \theta_A \circ \phi$ a flow minimizing the energy $\mathcal{E}^{\mathrm{diff}}$ subject to the constraints $\phi(0) = \mathbb{1}$ and $\theta_B = \theta_A \circ \phi(1)$ defines a geodesic path $(\theta(t))_{t\in[0,1]}$ with $\theta(t) = \theta_A \circ \phi^{-1}(t)$ in the spaces of images connecting θ_A and θ_B. If we aim at deriving a Riemannian distance directly between images via the flow of diffeomorphism approach, then a motion field v can be viewed as a representation of an image variation. Obviously, different motion fields might represent the same image variation. Hence, the corresponding equivalence class $\overline{v}$ is considered as a tangent vector on the image manifold and the associate metric is now given by

$$\mathcal{G}^{\mathrm{diff}}[\overline{v},\overline{v}] := \min_{v\in\overline{v}} \int_D L[v,v]\,\mathrm{d}x\,. \tag{4}$$

Consequently, the path energy on a path $(\theta(t))_{t\in[0,1]}$ reads as

$$\mathcal{E}^{\mathrm{diff}}[\theta] = \int_0^1 \mathcal{G}^{\mathrm{diff}}[\overline{v},\overline{v}]\,\mathrm{d}t\,, \tag{5}$$

which one minimizes over all image paths with $\theta(0) = \theta_A$, $\theta(1) = \theta_B$. Here, we assume that there is at least one path of finite path energy connecting θ_A and θ_B. In medical applications [BMTY02] each diffeomorphisms represents a particular anatomic configuration of an anatomic reference structures. For more details we refer to [DGM98a, BMTY05, JM00, MTY02].

Metamorphosis. A one-to-one correspondence of image grey values in warping applications is frequently not realistic. The metamorphosis approach offers a suitable generalization of the flow of diffeomorphism concept. It is first presented by Miller and Younes [MY01]. A rigorous analytical treatment is due to Trouvé and Younes [TY05]. In addition to the transport of image intensities along motion paths the variation of an intensity value along a motion path is allowed and reflected by an additional term in the energy. This term measures the integrated squared material derivative $z = \partial_t \theta + \nabla\theta \cdot v$. From a geometric perspective a pair of material derivative z and motion velocity v represent a variation of an image θ. Hence, an equivalence class $\overline{(z,v)}$ of all pairs which generate the same image variation is considered as a tangent vector on the image manifold. A Riemannian metric $\mathcal{G}^{\mathrm{meta}}[\overline{(z,v)},\overline{(z,v)}]$ acts on these tangent vectors and the associated path energy along an image path $(\theta(t))_{t\in[0,1]}$ is given by

$$\mathcal{E}^{\mathrm{meta}}[\theta] = \int_0^1 \mathcal{G}^{\mathrm{meta}}[\overline{(z,v)},\overline{(z,v)}]\,\mathrm{d}t\,. \tag{6}$$

As an example for the underlying Riemannian metric we obtain

$$\mathcal{G}^{\mathrm{meta}}[\overline{(z,v)},\overline{(z,v)}] = \min_{(z,v)\in\overline{(z,v)}} \int_D L[v,v] + \frac{1}{\delta}z^2\,\mathrm{d}x\,, \tag{7}$$

where the first three terms in the integrant retrieve the metric from the flow of diffeomorphism approach and encode the induced viscous dissipation, whereas the last term penalizes temporal changes of intensities along motion paths.

In this paper, we combine the optimal transportation approach with the metamorphosis approach. Thereby, in addition to the transportation cost we take into account a density variation of the transported measure and viscous dissipation. The paper is organized as follows: First we present our generalized image transport model in Section 1. In Section 2 we prove existence of geodesics in the non-viscous case. Then we propose a variational time discretization of the full model in Section 3, prove existence of time discrete geodesics in Section 4 and describe a fully discrete solution scheme in Section 5. Furthermore, we consider in Section 6 the algorithm used in [BB00] to compute for comparison reasons geodesics in the purely non-viscous case. Finally, we generalize in Section 7 our model to discrete weighted barycentres and apply it to textures and objects.

1   The generalized image transport model

In this section we will discuss the generalization of the optimal transport model in image warping and blending. These generalizations are motivated by two observations in applications:
– Frequently, objects or structures in images, which are in correspondence and are expected to be matched via the transport,

have different masses. From a global perspective the assumptions that images are considered as probability distributions is too restrictive. Indeed, the latter requires in advance contrast modulation, which is somewhat artificial. In the classical optimal transport model local mass differences lead to artifacts, where a local mass surplus has to be deposited elsewhere without any structural correspondence. We will no longer enforce the source free transport equation and explicitly incorporate a source term in the path energy which measures density modulation.

– Different from the flow of diffeomorphism approach the optimal transport maps are not necessarily homeomorphisms. On the other hand in many applications one is interested in topological consistency and at the same time the physical background of the application might suggest to incorporate a dissipative term in the path energy. Hence, we combine the classical transport cost model with a weighted viscous dissipation model.

To this end we first recall the formulation (2) of the Wasserstein distance proposed by Benamou and Brenier [BB00]. In what follows we restrict to a bounded domain $D \subset \mathbb{R}^d$ ($d = 2, 3$) with Lipschitz boundary. Now, we allow for a source term $z : [0, 1] \times D \to \mathbb{R}$ in the transport equation defined for given image intensity $\theta : [0, 1] \times D \to \mathbb{R}$ and transport field $v : [0, 1] \times D \to \mathbb{R}^d$ as

$$z := \partial_t \theta + \text{div}(\theta v) \,. \tag{8}$$

Furthermore, we pick up the model for the viscous dissipation in (5) and obtain as a new path energy

$$\mathcal{E}_{\delta,\gamma}[\theta] = \int_0^1 \mathcal{G}_{\delta,\gamma}[\overline{(z,v)}, \overline{(z,v)}] \, \mathrm{d}t \,. \tag{9}$$

with

$$\mathcal{G}_{\delta,\gamma}[\overline{(z,v)}, \overline{(z,v)}] = \min_{(z,v) \in \overline{(z,v)}} \int_D \theta|v|^2 + \frac{1}{\delta} z^2 + \gamma L[v,v] \, \mathrm{d}x \,, \tag{10}$$

which we minimize subject to (8) and the constraints $\theta(0) = \theta_A$ and $\theta(1) = \theta_B$. Here, θ_A and θ_B are the given input images and the $\overline{(z,v)}$ is the equivalence class of pairs of a source term and a transport field, which are consistent with the transport equation (8) for given image intensity θ. The involved local rate of viscous dissipation is given by $L[v,v] = \frac{\lambda}{2}(\text{tr}\varepsilon[v])^2 + \mu\text{tr}(\varepsilon[v]^2) + \epsilon|D^m v|^2$, where $\varepsilon[v] = \frac{1}{2}(\nabla v + \nabla v^T)$, $m > 1 + \frac{d}{2}$ and λ, μ, $\eta > 0$. (The first two terms represent the viscous dissipation of a Newtonian fluid and the higher order terms reflect a multipolar viscosity). As in [TY05] and similar to [AGS06] the condition $\partial_t \theta + \text{div}(\theta v) = z$ has to be understood in weak form

$$\int_0^1 \int_D \eta z \, \mathrm{d}x \, \mathrm{d}t = -\int_0^1 \int_D (\partial_t \eta + v \cdot \nabla \eta)\theta \, \mathrm{d}x \, \mathrm{d}t$$

for all $\eta \in C_c^\infty((0,1) \times D)$.

The first term in the metric is the classical transport cost rate, the second terms reflects the source term which measures the density modulation of the image intensity and the last term is the dissipation rate based on a multipolar viscous fluid model. Let us emphasize that for general non divergence free motion fields z does not coincide with the material derivative as in the metamorphosis model [TY05]. We suppose that $\gamma \geq 0$ measuring the impact of viscosity and $\delta > 0$ is a penalty parameter weighting the impact of density modulation on the metric and the path energy. In the formal limit $\delta \to \infty$ and for $\gamma = 0$ we retrieve the standard transport cost. Given the path energy, we can define a Riemannian (generalized Wasserstein) distance $\mathcal{W}_{\delta,\gamma}[\theta_A, \theta_B]$ of two images θ_A and θ_B as

$$\mathcal{W}_{\delta,\gamma}[\theta_A, \theta_B]^2 = \min_{\substack{(\theta(t))_{t \in [0,1]} \\ \theta(0)=\theta_A, \, \theta(1)=\theta_B}} \mathcal{E}_{\delta,\gamma}[\theta] \,. \tag{11}$$

2 Existence of geodesics for the non-viscous model ($\gamma = 0$)

In this section we study existence of minimizers θ of (9) in the non-viscous case, that is for $\gamma = 0$. In order to give a rigorous proof, it will be necessary to reformulate the formal problem (9) as a problem for measures rather than for densities. In particular, it will be crucial to treat the singular parts of the measures in an appropriate way.

Following [BB00], it will be useful to replace the velocity variable v by the momentum variable $w = \theta v$. Therefore, the function $\theta|v|^2$ appearing in the path energy (10) will be replaced by $|w|^2/\theta$. The joint convexity of this function will play a crucial role in the sequel.

The argument presented here is a modification of the argument in [DNS09] and our presentation follows this latter work very closely. Some additional arguments are needed to deal with the possibly varying total mass. On the other hand, some simplifications can be made, since we work on a bounded spatial domain instead of the whole space $\mathbb{R}^d$.

First we reformulate the action functional (10). Let Ω be a bounded domain in Euclidean space and fix a reference measure $\mathscr{L} \in \mathscr{M}^+(\Omega)$. In our application, the domain Ω will either be the spatial domain D or the space-time domain $[0,1] \times D$, and $\mathscr{L}$ will be the corresponding Lebesgue measure.

Let $\mu \in \mathscr{M}^+(\Omega)$, $\nu \in \mathscr{M}(\Omega; \mathbb{R}^d)$, and $\zeta \in \mathscr{M}(\Omega)$. The Lebesgue decomposition of these measures with respect to $\mathscr{L}$ is given by

$$\mu = \theta\mathscr{L} + \mu^\perp \,, \qquad \nu = w\mathscr{L} + \nu^\perp \,, \qquad \zeta = z\mathscr{L} + \zeta^\perp \,.$$

Let now $\mathscr{L}^\perp \in \mathscr{M}^+(\Omega)$ be such that $\mu^\perp, \nu^\perp$ and $\zeta^\perp$ are absolutely continuous with respect to $\mathscr{L}^\perp$ (take for instance $\mathscr{L}^\perp = \mu^\perp + |\nu^\perp| + |\zeta^\perp|$). Then we may write

$$\mu^\perp = \theta^\perp \mathscr{L}^\perp \,, \qquad \nu^\perp = w^\perp \mathscr{L}^\perp \,, \qquad \zeta^\perp = z^\perp \mathscr{L}^\perp \,.$$

As in the Benamou-Brenier formulation of the 2-Wasserstein distance we consider the function $\phi : [0, \infty) \times \mathbb{R}^d \to [0, \infty]$ defined by

$$\phi(\theta, w) = \begin{cases} 0 & \theta = 0 \text{ and } w = 0 \,, \\ \frac{|w|^2}{\theta} & \theta > 0 \,, \\ +\infty & \theta = 0 \text{ and } w \neq 0 \,. \end{cases}$$

Note that ϕ is lower-semicontinuous, convex and 1-homogeneous. The action functional $\mathcal{D} : \mathscr{M}^+(\Omega) \times \mathscr{M}(\Omega; \mathbb{R}^d) \times \mathscr{M}(\Omega) \to [0, +\infty]$ that we are interested in is given by

$$\mathcal{D}(\mu, \nu, \zeta) := \mathcal{D}_{\mathrm{BB}}(\mu, \nu) + \frac{1}{\delta}\mathcal{D}_Z(\zeta) \,, \qquad \mathcal{D}_{\mathrm{BB}}(\mu, \nu) := \int_\Omega \phi(\theta, w) \, \mathrm{d}\mathscr{L} + \int_\Omega \phi(\theta^\perp, w^\perp) \, \mathrm{d}\mathscr{L}^\perp \,,$$

$$\mathcal{D}_Z(\zeta) := \begin{cases} \int_\Omega z^2 \, \mathrm{d}\mathscr{L} & \zeta^\perp = 0 \,, \\ +\infty & \zeta^\perp \neq 0 \,. \end{cases}$$

Since ϕ is jointly 1-homogeneous, the definition of $\mathcal{D}_{\mathrm{BB}}$ does not depend on the choice of $\mathscr{L}^\perp$. The same is true for $\mathcal{D}_Z$, since we may write $\mathcal{D}_Z(\zeta) = \int_\Omega z^2 \, \mathrm{d}\mathscr{L} + \int_\Omega \psi(z^\perp) \, \mathrm{d}\mathscr{L}^\perp$, where $\psi : \mathbb{R} \to [0, +\infty]$ is the 1-homogeneous function defined by $\psi(0) = 0$ and $\psi(r) = +\infty$ for $r \neq 0$. Sometimes it will be useful to write $\mathcal{D}^\Omega$ instead of $\mathcal{D}$ in order to emphasize the domain Ω.

The following result is an immediate consequence of general lower-semicontinuity results for integral functionals on measures [AB88, AFP00].

Proposition 2.1 (Lower semicontinuity of the functional $\mathcal{D}$). *Consider weak*-convergent sequences of measures*

$$\mu_n \rightharpoonup^* \mu \in \mathscr{M}^+(\Omega) \,, \qquad \nu_n \rightharpoonup^* \nu \in \mathscr{M}(\Omega; \mathbb{R}^d) \,, \qquad \zeta_n \rightharpoonup^* \zeta \in \mathscr{M}(\Omega) \,.$$

Then we have $\mathcal{D}(\mu, \nu, \zeta) \leq \liminf_{n \to \infty} \mathcal{D}(\mu_n, \nu_n, \zeta_n)$.

Proof. The result follows, since both $\mathcal{D}_{\mathrm{BB}}$ and $\mathcal{D}_Z$ satisfy the assumptions of [DNS09, Theorem 2.1]. $\qquad \square$

The following crucial lemma is a special case of [DNS09, Proposition 3.6].

Lemma 2.2 (Integrability estimate). *Let $\mu \in \mathscr{M}^+(\Omega)$ and $\nu \in \mathscr{M}(\Omega; \mathbb{R}^d)$. For any Borel function $\eta : \Omega \to \mathbb{R}_+$ we have*

$$\int_\Omega \eta(x) \, \mathrm{d}|\nu|(x) \leq \left(\mathcal{D}_{BB}(\mu, \nu) \right)^{\frac{1}{2}} \left(\int_\Omega \eta^2 \, \mathrm{d}\mu \right)^{\frac{1}{2}} \,.$$

Proof. Set $S := \{x \in \Omega : \eta(x) > 0\}$. Using the scalar inequality $\sqrt{ab} + \sqrt{cd} \leq \sqrt{a+c}\sqrt{b+d}$ which holds for $a, b, c, d \geq 0$, we obtain

$$\int_S \eta(x) \, \mathrm{d}|\nu|(x) = \int_S \eta|w| \, \mathrm{d}\mathscr{L} + \int_S \eta|w^\perp| \, \mathrm{d}\mathscr{L}^\perp$$

$$\leq \left(\int_S \phi(\theta, w) \, \mathrm{d}\mathscr{L} \right)^{\frac{1}{2}} \left(\int_S \eta^2\theta \, \mathrm{d}\mathscr{L} \right)^{\frac{1}{2}} + \left(\int_S \phi(\theta^\perp, w^\perp) \, \mathrm{d}\mathscr{L}^\perp \right)^{\frac{1}{2}} \left(\int_S \eta^2\theta^\perp \, \mathrm{d}\mathscr{L}^\perp \right)^{\frac{1}{2}}$$

$$\leq \left(\int_S \phi(\theta, w) \, \mathrm{d}\mathscr{L} + \int_S \phi(\theta^\perp, w^\perp) \, \mathrm{d}\mathscr{L}^\perp \right)^{\frac{1}{2}} \left(\int_S \eta^2\theta \, \mathrm{d}\mathscr{L} + \int_S \eta^2\theta^\perp \, \mathrm{d}\mathscr{L}^\perp \right)^{\frac{1}{2}}$$

$$\leq \left(\mathcal{D}_{\mathrm{BB}}(\mu, \nu) \right)^{\frac{1}{2}} \left(\int_S \eta^2 \, \mathrm{d}\mu \right)^{\frac{1}{2}} \,.$$

$\square$

Let us now introduce the modified continuity equation.

Definition 2.3 (A continuity equation without conservation of mass)**.** *Let $t \mapsto \mu_t$ be weak*-continuous in $\mathscr{M}^+(D)$, let $t \mapsto \nu_t$ be Borel measurable in $\mathscr{M}(D; \mathbb{R}^d)$, and let $t \mapsto \zeta_t$ be Borel measurable in $\mathscr{M}(D)$. We say that the triple $(\mu_t, \nu_t, \zeta_t)_{t \in [0,1]}$ satisfies the continuity equation (and write $(\mu_t, \nu_t, \zeta_t)_{t \in [0,1]} \in \mathcal{CE}[0,1]$) if*

1. *the following integrability conditions hold:*

$$\int_0^1 |\nu_t|(D) \, \mathrm{d}t < \infty \,, \qquad \int_0^1 |\zeta_t|(D) \, \mathrm{d}t < \infty \,;$$

2. *the modified continuity equation $\partial_t \mu + \mathrm{div}(\nu) = \zeta$ holds in the sense of distributions,* i.e.*, for all space-time test functions $\eta \in C_0^1((0,1) \times \overline{D})$ we have*

$$\int_0^1 \int_D \partial_t \eta(t,x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t + \int_0^1 \int_D \nabla \eta(t,x) \cdot \mathrm{d}\nu_t(x) \, \mathrm{d}t + \int_0^1 \int_D \eta(t,x) \, \mathrm{d}\zeta_t(x) \, \mathrm{d}t = 0 \,.$$

A standard approximation argument (see [DNS09, Lemma 4.1]) shows that solutions to $\mathcal{CE}[0,1]$ satisfy, for all $0 \leq t_0 \leq t_1 \leq 1$,

$$\int_D \eta(t_1,x) d\mu_{t_1}(x) - \int_D \eta(t_0,x) d\mu_{t_0}(x) = \int_{t_0}^{t_1} \int_D \partial_t \eta(t,x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t + \int_{t_0}^{t_1} \int_D \nabla \eta(t,x) \cdot \mathrm{d}\nu_t(x) \, \mathrm{d}t$$
$$+ \int_{t_0}^{t_1} \int_D \eta(t,x) \, \mathrm{d}\zeta_t(x) \, \mathrm{d}t \tag{12}$$

for all space-time test functions $\eta \in C^1([0,1] \times \overline{D})$. In particular, taking $\eta(t,x) \equiv 1$, it follows that the increase of mass is given by

$$\mu_{t_1}(D) - \mu_{t_0}(D) = \int_{t_0}^{t_1} \zeta_t(D) \, \mathrm{d}t \,. \tag{13}$$

We are now in a position to rigorously define the extended distance $\mathcal{W}_\delta$ that was formally introduced in (11).

Definition 2.4. *For $\mu_A, \mu_B \in \mathscr{M}^+(D)$ we define $\mathcal{W}_\delta(\mu_A, \mu_B) \in [0, +\infty]$ by*

$$\mathcal{W}_\delta(\mu_A, \mu_B) := \inf_{\mu,\nu,\zeta} \left\{ \left(\int_0^1 \mathcal{D}(\mu_t, \nu_t, \zeta_t) \, \mathrm{d}t \right)^{1/2} : (\mu_t, \nu_t, \zeta_t)_{t \in [0,1]} \in \mathcal{CE}[0,1] \,, \ \mu_0 = \mu_A \,, \ \mu_1 = \mu_B \right\}. \tag{14}$$

The following theorem is the main result of this section.

Theorem 2.5 (Existence of geodesics)**.** *Let $\delta \in (0, \infty)$ and take $\mu_A, \mu_B \in \mathscr{M}^+(D)$ with $\mathcal{W}_\delta(\mu_A, \mu_B) < \infty$. Then there exists a minimizer $(\overline{\mu}_t, \overline{\nu}_t, \overline{\zeta}_t)_{t \in [0,1]}$ that realizes the infimum in (14). Moreover, the associated curve $(\overline{\mu}_t)_{t \in [0,1]}$ is a constant speed geodesic for $\mathcal{W}_\delta$,* i.e.,

$$\mathcal{W}_\delta(\overline{\mu}_s, \overline{\mu}_t) = |s - t| \mathcal{W}_\delta(\mu_A, \mu_B)$$

for all $s, t \in [0,1]$. Furthermore, we have the alternative characterisation

$$\mathcal{W}_\delta(\mu_A, \mu_B) := \inf_{\mu,\nu,\zeta} \left\{ \int_0^1 \sqrt{\mathcal{D}(\mu_t, \nu_t, \zeta_t)} \, \mathrm{d}t : (\mu_t, \nu_t, \zeta_t)_{t \in [0,1]} \in \mathcal{CE}[0,1] \,, \ \mu_0 = \mu_A \,, \ \mu_1 = \mu_B \right\}.$$

Proof. The existence of a minimizer is an immediate consequence of Proposition 2.6 below. The remaining statements follow by standard arguments, see [DNS09, Theorem 5.4] for details. $\qquad\square$

Let us now state and prove the main ingredient for the proof of Theorem 2.5. We write $\int_0^1 \delta_t \otimes \mu_t \, \mathrm{d}t$ to denote the measure μ on $[0,1] \times D$ satisfying

$$\int_{[0,1] \times D} \eta(t,x) \, \mathrm{d}\mu(t,x) = \int_0^1 \int_D \eta(t,x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t$$

for all $\eta \in C([0,1] \times D)$.

Proposition 2.6 (Compactness for solutions to the continuity equation with bounded action). *Suppose that* $(\mu_t^n, \nu_t^n, \zeta_t^n)_{t\in(0,1)} \in \mathcal{CE}[0,1]$ *satisfy*

(A1) $M_1 := \sup_n \mu_0^n(D) < \infty$;

(A2) $M_2 := \sup_n \int_0^1 \mathcal{D}(\mu_t^n, \nu_t^n, \zeta_t^n)\,\mathrm{d}t < \infty$.

Set $\nu^n := \int_0^1 \delta_t \otimes \nu_t^n\,\mathrm{d}t \in \mathscr{M}([0,1]\times D; \mathbb{R}^d)$ *and* $\zeta^n := \int_0^1 \delta_t \otimes \zeta_t^n\,\mathrm{d}t \in \mathscr{M}([0,1]\times D)$. *Then, there exists a subsequence (again indexed by n) and a triple* $(\mu_t, \nu_t, \zeta_t)_{t\in(0,1)} \in \mathcal{CE}[0,1]$ *such that*

1. $\mu_t^n \rightharpoonup^* \mu_t$ *in* $\mathscr{M}^+(D)$ *for all* $t \in [0,1]$;

2. $\nu^n \rightharpoonup^* \nu$ *in* $\mathscr{M}([0,1]\times D; \mathbb{R}^d)$;

3. $\zeta^n \rightharpoonup^* \zeta$ *in* $\mathscr{M}([0,1]\times D)$.

Moreover, for the above subsequence

$$\int_0^1 \mathcal{D}(\mu_t, \nu_t, \zeta_t)\,\mathrm{d}t \le \liminf_{n\to\infty} \int_0^1 \mathcal{D}(\mu_t^n, \nu_t^n, \zeta_t^n)\,\mathrm{d}t\,. \tag{15}$$

Proof. In view of (A2), we first observe that

$$M_3 := \sup_n \int_0^1 |\zeta_t^n|(D)\,\mathrm{d}t = \sup_n \int_0^1 \int_D |z_t^n(x)|\,\mathrm{d}x\,\mathrm{d}t < \infty\,.$$

Therefore, (13) yields the uniform bound

$$\mu_t^n(D) \le \mu_0^n(D) + \int_0^t |\zeta_s^n|(D)\,\mathrm{d}s \le M_1 + M_3$$

for all n. Moreover, Lemma 2.2 implies that

$$|\nu_t^n|(D) \le \sqrt{\mu_t^n(D)\mathcal{D}_{\mathrm{BB}}(\mu_t^n, \nu_t^n)}\,,$$

hence by the Hölder inequality we obtain

$$\int_0^1 |\nu_t^n|(D)^2\,\mathrm{d}t \le \int_0^1 \mu_t^n(D)\mathcal{D}(\mu_t^n, \nu_t^n, \zeta_t^n)\,\mathrm{d}t \le M_2(M_1 + M_3)\,,$$

which shows that the maps $\{t \mapsto |\nu_t^n|(D)\}_n$ are uniformly bounded in $L^2(0,1)$, hence uniformly integrable.

Since $|\nu^n|([0,1]\times D) \le (\int_0^1 |\nu_t^n|(D)^2\,\mathrm{d}t)^{\frac{1}{2}}$, the measures $\{\nu^n\}_n \in \mathscr{M}([0,1]\times D; \mathbb{R}^d)$ have uniformly bounded total variation on $[0,1]\times D$, hence we can extract a subsequence that converges weakly* to some measure $\nu \in \mathscr{M}([0,1]\times D; \mathbb{R}^d)$. The uniform integrability of $\{t \mapsto |\nu_t^n|(D)\}_n$ implies that the image measure of ν under the mapping $(t,x) \mapsto t$ is absolutely continuous with respect to the Lebesgue measure on $[0,1]$. Therefore, the disintegration theorem (see, *e.g.*, [AGS06, Theorem 5.3.1]) allows us to write $\nu = \int_0^1 \delta_t \otimes \nu_t\,\mathrm{d}t$ for some family of measures $\{\nu_t\}_{t\in[0,1]} \in \mathscr{M}(D; \mathbb{R}^d)$.

Fix $0 \le \tau \le 1$, take $\eta \in C^1(D)$, and set $\bar\xi(t,x) := \nabla\eta(x)\chi_{[0,\tau]}(t)$. Although $\bar\xi$ is discontinuous, general approximation results (see [AGS06, Proposition 5.1.10]) imply that

$$\int_0^\tau \int_D \nabla\eta(x)\,\mathrm{d}\nu_t^n(x)\,\mathrm{d}t = \int_{[0,1]\times D} \bar\xi(t,x)\,\mathrm{d}\nu^n(t,x) \to \int_{[0,1]\times D} \bar\xi(t,x)\,\mathrm{d}\nu(t,x) = \int_0^\tau \int_D \nabla\eta(x)\,\mathrm{d}\nu_t(x)\,\mathrm{d}t\,. \tag{16}$$

Let us now consider the term involving ζ_t^n, which is treated similarly. For all n and a.e. $t \in [0,1]$ we use (A2) to conclude that $\zeta_t^n = z_t^n\mathscr{L}$. Therefore we obtain

$$\int_0^1 |\zeta_t^n|(D)^2\,\mathrm{d}t = \int_0^1 \left(\int_D |z_t^n(x)|\,\mathrm{d}x \right)^2\,\mathrm{d}t < \infty\,.$$

As above, we infer that the mappings $\{t \mapsto |\zeta_t^n|(D)\}_n$ are uniformly integrable, and that there exists a subsequence of $\{\zeta^n\}_n$ that convergence weakly* to some measure $\zeta \in \mathscr{M}([0,1]\times D)$. By the disintegration theorem we may write

$\zeta = \int_0^1 \delta_t \otimes \zeta_t \, \mathrm{d}t$ for a family of measures $\{\zeta_t\}_{t \in [0,1]} \in \mathscr{M}(D)$. Set $\tilde{\xi}(t, x) := \eta(x)\chi_{[0,\tau]}(t)$. Arguing as above, we obtain

$$\int_0^\tau \int_D \eta(x) \, \mathrm{d}\zeta_t^n(x) \, \mathrm{d}t = \int_{[0,1] \times D} \tilde{\xi}(t, x) \, \mathrm{d}\zeta^n(t, x) \to \int_{[0,1] \times D} \tilde{\xi}(t, x) \, \mathrm{d}\zeta(t, x) = \int_0^\tau \int_D \eta(x) \, \mathrm{d}\zeta_t(x) \, \mathrm{d}t \, . \qquad (17)$$

We are now in a position to obtain subsequential convergence of $\{\mu_t^n\}_n$. Indeed, it follows from (12) that

$$\int_D \eta(x) d\mu_\tau^n(x) = \int_D \eta(x) d\mu_0^n(x) + \int_0^\tau \int_D \nabla \eta(x) \cdot \mathrm{d}\nu_t^n(x) \, \mathrm{d}t + \int_0^\tau \int_D \eta(x) \, \mathrm{d}\zeta_t^n(x) \, \mathrm{d}t \, .$$

Moreover, (A1) implies that there exists a measure $\mu_0 \in \mathscr{M}^+(D)$ such that $\mu_0^n \rightharpoonup^* \mu_0$ (after passing to a subsequence). In view of (16) and (17) the latter equation implies weak*-convergence of $\{\mu_\tau^n\}_n$ to some measure μ_τ for every $\tau \in [0, 1]$. It is readily checked that $(\mu_t, \nu_t, \zeta_t)_t \in \mathcal{CE}[0, 1]$.

It remains to prove (15). For this purpose we write $\mu = \int_0^1 \delta_t \otimes \mu_t \, \mathrm{d}t$ and $\mu^n = \int_0^1 \delta_t \otimes \mu_t^n \, \mathrm{d}t$. It is straightforward to check that μ^n converges weakly* to μ in $\mathscr{M}_+([0, 1] \times D)$. Now the result follows by observing that

$$\int_0^1 \mathcal{D}^D(\mu_t, \nu_t, \zeta_t) \, \mathrm{d}t = \mathcal{D}^{[0,1] \times D}(\mu, \nu, \zeta) \, ,$$

and applying Proposition 2.1 to $\mathcal{D}^{[0,1] \times D}$. □

3   A variational time discretization

In what follows, we derive a time discrete approximation of the energy (9) and thereby a variational approach for the definition of geodesic paths. We refer to [WBRS11] for the general concept and to [RW14] for the numerical analysis in the context of shape spaces which are Hilbert manifolds and in [BER15] a variational time discretization of geodesics in the metamorphosis model is discussed.

As a motivation let us briefly present a toy model in finite dimensions. On a smooth m-dimensional manifold $\mathcal{M}$ embedded in $\mathbb{R}^d$ ($m \leq d$) we consider the simple energy $\mathbf{F}[y, \tilde{y}] = |\tilde{y} - y|^2$ which reflects the stored elastic energy in a spring spanned between points y and $\tilde{y}$ through the ambient space of $\mathcal{M}$ in $\mathbb{R}^d$. The smoothness of $\mathcal{M}$ implies that $\mathbf{F}[y, \tilde{y}] = \mathrm{dist}_{\mathcal{M}}(y, \tilde{y})^2 + O(\mathrm{dist}_{\mathcal{M}}(y, \tilde{y})^3)$, where $\mathrm{dist}_{\mathcal{M}}(y, \tilde{y})$ denotes the Riemannian distance between y and $\tilde{y}$. Hence, we can approximate the path length of a smooth path $(y(t))_{t \in [0,1]}$ via sampling $y_k = y(\frac{k}{K})$ and then evaluate the discrete path energy

$$\mathbf{E}^K[y_0, \ldots, y_K] = K \sum_{k=1}^K |y_k - y_{k-1}|^2 \, ,$$

such that $\mathbf{E}^K[y_0, \ldots, y_K]$ converges to $\mathcal{E}[(y(t))_{t \in [0,1]}] = \int_0^1 |\dot{y}(t)|^2 \, \mathrm{d}t$ for $K \to \infty$. Here, we use that $\frac{y_k - y_{k-1}}{\tau}$ is an approximation of the velocity $\dot{y}(k/K)$ where $\tau = \frac{1}{K}$ is the time step size of our discretization on the time interval $[0, 1]$. In fact, based on the approximation $\mathbf{F}$ of the squared distance $\mathrm{dist}_{\mathcal{M}}$, which is easy to implement, we obtain an effective approximation of the Riemannian path energy $\mathcal{E}$. Correspondingly, we call a minimizer $(y_0, \ldots, y_K)$ of the discrete path energy for fixed y_0 and y_K a discrete geodesic. We refer to [RW14, BER15] for a detailed discussion, why the discrete path energy instead of the discrete path length is the right concept to compute discrete geodesics. In particular, Γ-convergence of the discrete path energy is proven in case of the metamorphosis model in [BER15] and under suitable assumptions in the context of Hilbert manifolds in [RW14].

Now, we ask for a similar time discrete approximation of the continuous path energy $\mathcal{E}_{\delta,\gamma}$ defined in (9). To this end, we consider a discrete path $(\theta_0, \ldots, \theta_K)$ in the space of image intensities with $\theta_k \in \mathcal{I}$ for $k = 1, \ldots, K$ with $\mathcal{I} := L^2(D, \mathbb{R}_{\geq 0})$ and ask for a matching functional $\mathbf{F}$ on consecutive pairs $\theta, \tilde{\theta}$ of image intensities. In fact, this matching functional should reflect time discrete counterparts of all three ingredients of the metric $\mathcal{G}_{\delta,\gamma}$ and the induced continuous path energy $\mathcal{E}_{\delta,\gamma}$, namely the transport cost, the viscous dissipation and the source term. Like in the original Monge problem we take into account deformations ϕ in a suitable space $\mathcal{A}$ of admissible deformations, to be defined later, and optimize for given $\theta, \tilde{\theta}$ a suitable functional $\mathbf{F}^D[\theta, \tilde{\theta}, \cdot]$ over all admissible deformations to define the value of the matching functional $\mathbf{F}[\theta, \tilde{\theta}]$, i.e.

$$\mathbf{F}[\theta, \tilde{\theta}] = \inf_{\phi \in \mathcal{A}} \mathbf{F}^D[\theta, \tilde{\theta}, \phi] \, .$$

With the matching functional at hand we then define the discrete path energy summing over applications of the matching functional to consecutive pairs of image intensities θ_{k-1} and θ_k of a discrete path $(\theta_0, \ldots, \theta_K)$ and get

$$\mathbf{E}_{\delta,\gamma}^K[\theta_0, \ldots, \theta_K] = K \sum_{k=1}^K \mathbf{F}[\theta_{k-1}, \theta_k]. \tag{18}$$

Thus, the resulting time discrete approximation of the squared Riemannian distance is given by

$$\mathcal{W}_{\delta,\gamma}^K[\theta_A, \theta_B]^2 = \min_{\substack{\theta_0,\ldots,\theta_K \in \mathcal{I} \\ \theta_0 = \theta_A, \, \theta_K = \theta_B}} \mathbf{E}_{\delta,\gamma}^K[\theta_0, \ldots, \theta_K]. \tag{19}$$

Here, we assume $\theta_A, \theta_B \in \mathcal{I}$. In what follows we list now the appropriate components of $\mathbf{F}$ reflecting the different ingredients of the continuous path energy.

Approximation of the transport cost. To approximate the first term in the metric (10) we make use of the equivalence of the original Monge problem and the Benamou Brenier formulation [BB00] of optimal transport and define

$$\mathbf{F}_{\text{transport}}^D[\theta, \phi] = \int_D |\phi - 1\!\!1|^2 \theta \, \mathrm{d}x. \tag{20}$$

Here, $\frac{\phi - 1\!\!1}{\tau}$ is an approximation of the transport velocity with $1\!\!1$ being the identity deformation.

Approximation of the density modulation cost. For a diffeomorphism ϕ the push forward condition $\phi_\#(\theta\mathscr{L}) = \tilde\theta\mathscr{L}$ can be expressed as

$$\theta = \det(D\phi)\tilde\theta \circ \phi.$$

As an approximation of the source term $z = \partial_t \theta + \mathrm{div}(v\theta)$ we take into account

$$\mathbf{F}_{\text{source}}^D[\theta, \tilde\theta, \phi] = \int_D \frac{1}{\delta} |\det(D\phi)\tilde\theta \circ \phi - \theta|^2 \, \mathrm{d}x.$$

Approximation of the dissipation cost. By Rayleigh's paradigm [Str45] one derives models for viscous dissipation from elastic energies replacing elastic strains by strain rates. We proceed as in [BER15], where a time discretization of the metamorphosis model was investigated and define

$$\mathbf{F}_{\text{viscous}}^D[\phi] = \gamma \int_D W(D\phi) + \epsilon |D^m \phi|^2 \, \mathrm{d}x.$$

Here, W is a hyper elastic energy density and the higher order term $|D^m\phi|^2$ acts as a regularizing term for some small $\epsilon > 0$ and enforces the deformations to be in the space $W^{m,2}$. We make the following assumptions on W (cf. also [BER15]):

(W1) W is non-negative and polyconvex,

(W2) $W(A) \geq \alpha_0 |\log \det A| - \alpha_1$ for $\alpha_0, \alpha_1 > 0$, and every invertible matrix A with $\det A > 0$, $W(A) = \infty$ for $\det A \leq 0$, and

(W3) W is sufficiently smooth and the following consistency assumptions with respect to the differential operator L hold true: $W(1\!\!1) = 0$, $DW(1\!\!1) = 0$ and $\frac{1}{2}D^2 W(1\!\!1)(B, B) = \frac{\lambda}{2}(\mathrm{tr}B)^2 + \mu\mathrm{tr}((\frac{B+B^T}{2})^2)$ for all $B \in \mathbb{R}^{d,d}$.

Due to the incorporation of this dissipation energy we finally define the space of admissible deformations over which we minimize in the definition of $\mathbf{F}[\theta, \tilde\theta]$ as

$$\mathcal{A} = \left\{ \phi \in W^{m,2}(D, D) \, : \, \det(D\phi) > 0 \text{ a.e. in } D, \phi = 1\!\!1 \text{ on } \partial D \right\},$$

We assume that $m > 1 + \frac{d}{2}$, which implies by Sobolev embedding that the admissible deformations are diffeomorphisms. Given these energy contributions we can define the compound energy

$$\mathbf{F}^D[\theta, \tilde\theta, \phi] = \mathbf{F}_{\text{transport}}^D[\theta, \phi] + \mathbf{F}_{\text{source}}^D[\theta, \tilde\theta, \phi] + \mathbf{F}_{\text{viscous}}^D[\phi].$$

The following interpolation results justifies our choice of the time discrete path energy.

Theorem 3.1 (Consistency of the discrete path energy). *For a convex domain D and a sufficiently smooth path of image intensities $(\theta(t))_{t\in[0,1]}$ with $\theta \geq 0$ a.e. in $[0,1] \times D$ and a sufficiently smooth family of velocities $(v(t))_{t\in[0,1]}$ we consider interpolated images $\theta_k^K = \theta(\frac{k}{K})$ and motion fields $v_k^K = v(\frac{k}{K})$. Then the resulting extended path energy*

$$\mathbf{E}_{\delta,\gamma}^{K,D}[\theta_0^K,\dots,\theta_K^K,\phi_1^K,\dots,\phi_K^K] := \sum_{k=1}^{K} \mathbf{F}^D[\theta_{k-1}^K, \theta_k^K, \phi_k^K] \tag{21}$$

with $\phi_k^K = \frac{1}{K}v_k^K + \mathbb{1}$ converges to the corresponding continuous path energy

$$\mathcal{E}_{\delta,\gamma}^D[\theta,v] := \int_0^1 \int_D \theta|v|^2 + \frac{1}{\delta}z^2 + \gamma L[v,v]\,\mathrm{d}x\,\mathrm{d}t\,.$$

Proof. We define the step size $\tau = \frac{1}{K}$. First, for the transport cost we easily get

$$K\sum_{k=1}^{K}\int_D |\phi_k^K - \mathbb{1}|^2\theta_{k-1}^K\,\mathrm{d}x = \sum_{k=1}^{K}\tau\int_D |v_k^K|^2\theta_{k-1}^K\,\mathrm{d}x \xrightarrow{K\to\infty} \int_0^1\int_D |v|^2\theta\,\mathrm{d}x\,\mathrm{d}t\,.$$

Following [WBRS11] the convergence of the dissipation cost follows from a Taylor expansion of the hyperelastic density function W by using the consistency assumptions:

$$K\sum_{k=1}^{K}\int_D W(D\phi_k^K) + \epsilon|D^m\phi_k^K|^2\,\mathrm{d}x$$

$$= \frac{1}{\tau}\sum_{k=1}^{K}\int_D W(\mathbb{1}) + \tau DW(\mathbb{1})(Dv_k^K) + \frac{\tau^2}{2}D^2W(\mathbb{1})(Dv_k^K,Dv_k^K) + O(\tau^3) + \epsilon\tau^2|D^m v_k^K|^2\,\mathrm{d}x$$

$$= \sum_{k=1}^{K}\tau\int_D \left(\frac{\lambda}{2}\left(\mathrm{tr}Dv_k^K\right)^2 + \mu\,\mathrm{tr}\left(\frac{Dv_k^K + (Dv_k^K)^T}{2}\right)^2\right) + O(\tau) + \epsilon|D^m v_k^K|^2\,\mathrm{d}x \xrightarrow{K\to\infty} \int_0^1\int_D L[v,v]\,\mathrm{d}x\,\mathrm{d}t\,.$$

Finally, for the density modulation cost we use the Taylor expansions

$$\det D\phi_k^K = \mathbb{1} + \tau\,\mathrm{tr}\left(\frac{D\phi_k^K - \mathbb{1}}{\tau}\right) + O(\tau^2) = \mathbb{1} + \tau\,\mathrm{div}(v_k^K) + O(\tau^2)$$

$$\theta_k^K \circ \phi_k^K = \theta_k^K + \tau\nabla\theta_k^K \cdot v_k^K + O(\tau^2)$$

and obtain

$$K\sum_{k=1}^{K}\int_D |\det(D\phi_k^K)\theta_k^K \circ \phi_k^K - \theta_{k-1}^K|^2\,\mathrm{d}x$$

$$= \sum_{k=1}^{K}\tau\int_D \left|\frac{\theta_k^K - \theta_{k-1}^K}{\tau} + \mathrm{div}(v_k^K)\theta_k^K + \nabla\theta_k^K \cdot v_k^K + O(\tau)\right|^2\,\mathrm{d}x \xrightarrow{K\to\infty} \int_0^1\int_D |\partial_t\theta + \mathrm{div}(\theta v)|^2\,\mathrm{d}x\,\mathrm{d}t\,.$$

$\square$

4 Existence of time discrete geodesics

In this section we assume that the assumptions of Section 3 are fulfilled and that $\delta, \gamma > 0$. As before we assume that $m > 1 + \frac{d}{2}$. We will show that for given images $\theta_A, \theta_B \in \mathcal{I} = L^2(D, \mathbb{R}_{\geq 0})$ a time discrete geodesic exists. First we prove that $\bar{\mathbf{F}}$ is well-posed in the sense that there is an optimal deformation between two images.

Proposition 4.1 (Existence of minimizing deformations). *Let $\theta, \tilde{\theta} \in \mathcal{I}$. Then $\mathbf{F}^D[\theta, \tilde{\theta}, \phi]$ attains its minimum over all deformation $\phi \in \mathcal{A}$. Moreover, ϕ is a diffeomorphism and $\phi^{-1} \in C^{1,\alpha}(D)$ for $\alpha \in (0, m-1-\frac{d}{2})$.*

Proof. *Step 1.* First, we observe that $\mathbf{F}^D$ is bounded from below, since θ is non-negative by definition of $\mathcal{I}$, $\mathcal{W}$ is non-negative by assumption (W1) and the source term is non-negative anyway. Because of (W2) and $\phi = 1\!\!1 \in \mathcal{A}$ there exists an upper bound for the energy $\mathbf{F}^D$ on a minimizing sequence $(\phi^j)_{j \in \mathbb{N}}$. Following [BER15] one observes that a subsequence, again denoted by (ϕ^j), converges weakly in $W^{m,2}(D, D)$ to some $\phi \in W^{m,2}(D, D)$ and for the limit deformation we get $\phi^{-1} \in C^{1,\alpha}(D)$.

Step 2. We prove that $\phi \mapsto \mathbf{F}^D[\theta, \tilde{\theta}, \phi]$ is lower semicontinuous w.r.t. weak convergence in L^2. It is sufficient to show that $\det(D\phi^j)\tilde{\theta} \circ \phi^j \rightharpoonup \det(D\phi)\tilde{\theta} \circ \phi$ in L^2. Then the result follows from the weak lower semicontinuity of the L^2-norm, the compact embedding of $W^{m,2}(D, D)$ into $C^{1,\alpha}(D, D)$ for $0 < \alpha < m - \frac{d}{2}$, and results on the weak lower semicontinuity of polyconvex functionals [Cia88]. By the assumption $\tilde{\theta} \in L^2(D)$ and by Step 1 we have a uniform L^2-bound on $\det(D\phi^j)\tilde{\theta} \circ \phi^j$, so it is enough to prove that the expression converges in the sense of distributions. For $\eta \in C_c^\infty(D)$ we have

$$\int_D \det(D\phi^j)(x)\tilde{\theta} \circ \phi^j(x)\eta(x)\,\mathrm{d}x = \int_D \tilde{\theta}(x)\eta \circ (\phi^j)^{-1}(x)\,\mathrm{d}x \quad \text{and}$$

$$\int_D \det(D\phi)(x)\tilde{\theta} \circ \phi(x)\eta(x)\,\mathrm{d}x = \int_D \tilde{\theta}(x)\eta \circ \phi^{-1}(x)\,\mathrm{d}x\,.$$

Since $(\phi^j)^{-1} \to \phi^{-1}$ in $C^{1,\alpha}$, the result follows by the dominated convergence theorem. $\qquad\square$

Now, for a given discrete path $(\theta_0, \ldots, \theta_K) \in \mathcal{I}^{K+1}$ we consider $\mathbf{E}_{\delta,\gamma}^{K,D}$ defined in (21). By Proposition 4.1 there exists $\Phi = (\phi_1, \ldots, \phi_K) \in \mathcal{A}^K$ such that $\mathbf{E}_{\delta,\gamma}^{K,D}[(\theta_0, \ldots, \theta_K), (\phi_1, \ldots, \phi_K)] = \mathbf{E}_{\delta,\gamma}^K[(\theta_0, \ldots, \theta_K)]$. Now we study $\mathbf{E}_{\delta,\gamma}^{K,D}$ for a fixed vector of deformations.

Proposition 4.2. *Let $\theta_A, \theta_B \in \mathcal{I}$, $K \geq 2$. Then for a fixed vector of deformations $\Phi = (\phi_1, \ldots, \phi_K) \in \mathcal{A}^K$ there exists a unique discrete path $(\theta_0, \ldots, \theta_K) \in \mathcal{I}^{K+1}$ with $\theta_0 = \theta_A$ and $\theta_K = \theta_B$, i.e.*

$$\mathbf{E}_{\delta,\gamma}^{K,D}[(\theta_0, \ldots, \theta_K), \Phi] = \inf_{(\theta_1, \ldots, \theta_{K-1}) \in \mathcal{I}^{K-1}} \mathbf{E}_{\delta,\gamma}^{K,D}[(\theta_A, \theta_1, \ldots, \theta_{K-1}, \theta_B), \Phi]\,.$$

Proof. First we see that the functional is bounded from above on a minimizing sequence by computing the energy of $(\theta_B, \ldots, \theta_B) \in \mathcal{I}^{K-1}$:

$$\mathbf{E}_{\delta,\gamma}^{K,D}[(\theta_A, \theta_B, \ldots, \theta_B, \theta_B), (\phi_1, \ldots, \phi_K)] \leq C(\phi_1, \ldots, \phi_k)(1 + \|\theta_A\|_2 + \|\theta_A\|_2^2 + \|\theta_B\|_2 + \|\theta_B\|_2^2) < \infty\,.$$

Next we observe that for fixed $(\phi_1, \ldots, \phi_K) \in \mathcal{A}^K$ the time discrete path energy is quadratically growing, *i.e.*

$$\mathbf{E}_{\delta,\gamma}^{K,D}[(\theta_A, \theta_1 \ldots, \theta_{K-1}, \theta_B), (\phi_1, \ldots, \phi_K)] \geq c_1 \sum_{i=1,\ldots,K-1} \|\theta_k\|_{L^2(D)}^2 - c_2$$

for constants $c_1, c_2 > 0$ depending on $(\phi_k)_k$. Therefore we can take a minimizing sequence $(\theta_1^j, \ldots, \theta_{K-1}^j)_{j \in \mathbb{N}} \subset \mathcal{I}^{K-1}$, which has because of the upper bound a weakly converging subsequence in L^2 with limit $(\theta_1, \ldots, \theta_{K-1}) \in \mathcal{I}^{K-1}$. Now, the energy is strictly convex in θ_k for all $k = 1, \ldots, K-1$, hence there is a unique minimizer in $\mathcal{I}^{K-1}$. $\qquad\square$

Next, we can use these two propositions to prove existence of minimizers of the discrete path energy $\mathbf{E}_{\delta,\gamma}^K$.

Theorem 4.3 (Existence of discrete geodesics)**.** *Let $\theta_A, \theta_B \in \mathcal{I}$, $K \geq 2$ be given. Then there exists $(\theta_1, \ldots, \theta_{K-1}) \in \mathcal{I}^{K-1}$ s.t.*

$$\mathbf{E}_{\delta,\gamma}^K[(\theta_A, \theta_1, \ldots, \theta_{K-1}, \theta_B)] = \inf_{(\tilde{\theta}_1, \ldots, \tilde{\theta}_{K-1}) \in \mathcal{I}^{K-1}} \mathbf{E}_{\delta,\gamma}^K[(\theta_A, \tilde{\theta}_1, \ldots, \tilde{\theta}_{K-1}, \theta_B)]\,.$$

Proof. Taking $\theta_k^j = \theta_B$ and $\phi_k = 1\!\!1$ to test the energy $\mathbf{E}_{\delta,\gamma}^K$ we observe that the $\mathbf{E}_{\delta,\gamma}^K$ is bounded from above on a minimizing sequence $(\theta_1^j, \ldots, \theta_{K-1}^j)_{j \in \mathbb{N}}$. Take a minimizing sequence $(\theta_1^j, \ldots, \theta_{K-1}^j)_{j \in \mathbb{N}}$ of the discrete path energy $\mathbf{E}_{\delta,\gamma}^K[(\theta_A, \cdot, \theta_B)]$. Due to Proposition 4.1, for every $(\theta_1^j, \ldots, \theta_{K-1}^j)$ there exists a family of optimal deformations $(\phi_1^j, \ldots, \phi_K^j) \in \mathcal{A}^K$ with $\mathbf{E}_{\delta,\gamma}^{K,D}[(\theta_A, \theta_1^j, \ldots, \theta_{K-1}^j, \theta_B), (\phi_1^j, \ldots, \phi_K^j)] \leq \mathbf{E}_{\delta,\gamma}^{K,D}[(\theta_A, \theta_1^j, \ldots, \theta_{K-1}^j, \theta_B), \Psi]$ for all $\Psi \in \mathcal{A}^K$. As in the proof of Proposition 4.1 there exists a subsequence again denoted $(\phi_k^j)_{j \in \mathbb{N}}$ with $\phi_k^j \rightharpoonup \phi_k$ in $W^{m,2}$ for all $k = 1, \ldots, K$, s.t. $\phi_k^{-1} \in C^{1,\alpha}$. By Proposition 4.2 we can assume (possible replacing $(\theta_1^j, \ldots, \theta_{K-1}^j)$ and thereby further reducing the energy) that $(\theta_1^j, \ldots, \theta_{K-1}^j)$ already minimizes the energy $\mathbf{E}_{\delta,\gamma}^{K,D}[(\theta_A, \cdot, \theta_B), (\phi_1^j, \ldots, \phi_K^j)]$ in $\mathcal{I}^{K-1}$.

11

Then θ_k^j is uniformly bounded in L^2 by a constant C depending only on θ_A and θ_B for $k = 1, \ldots, K$. This constant C is independent of the ϕ_k^j due to the uniform bound of ϕ_k^j in $W^{m,2}$. Hence we can pass to a further subsequence satisfying $(\theta_1^j, \ldots, \theta_{K-1}^j) \rightharpoonup (\theta_1, \ldots, \theta_{K-1})$ in L^2. To prove weak lower semicontinuity in L^2 of the functional, it is sufficient to pass to the limit in the identities

$$\int_D \det(D\phi_k^j)(x)\theta_k^j \circ \phi_k^j(x)\eta(x)\,\mathrm{d}x = \int_D \theta_k^j(x)\eta \circ (\phi_k^j)^{-1}(x)\,\mathrm{d}x\,, \tag{22}$$

$$\int_D \det(D\phi_k)(x)\theta_k \circ \phi_k(x)\eta(x)\,\mathrm{d}x = \int_D \theta_k(x)\eta \circ \phi_k^{-1}(x)\,\mathrm{d}x\,, \tag{23}$$

for an arbitrary C^∞-function η, which follows from the $C^{1,\alpha}$-convergence of $(\phi_k^j)^{-1}$ and the weak L^2-convergence of θ_k^j. For the demonstration of lower semicontinuity in the remaining terms we refer to analogous discussion in Proposition 4.1. $\qquad\square$

Finally, let us study in more detail the optimality conditions for $(\theta_1, \ldots, \theta_{K-1}) \in \mathcal{I}^{K-1}$ in preparation of the later derivation of a numerical algorithm. At first we consider the simplified model without the constraint $\theta_k \geq 0$ for $k = 1, \ldots, K-1$. Since for fixed deformations the energy is strictly convex, there exists a unique minimizer. For each $k = 1, \ldots, K-1$ there are two terms in the energy where θ_k appears:

$$\mathbf{F}^D[\theta_k, \theta_{k+1}, \phi_{k+1}] = \int_D |\phi_{k+1} - \mathrm{I\!I}|^2\theta_k + \frac{1}{\delta}|\det(D\phi_{k+1})\theta_{k+1} \circ \phi_{k+1} - \theta_k|^2\,\mathrm{d}x + \gamma\mathbf{F}^D_{\text{viscous}}[\phi_{k+1}]\,,$$

$$\mathbf{F}^D[\theta_{k-1}, \theta_k, \phi_k] = \int_D |\phi_k - \mathrm{I\!I}|^2\theta_{k-1} + \frac{1}{\delta}|\det(D\phi_k)\theta_k \circ \phi_k - \theta_{k-1}|^2\,\mathrm{d}x + \gamma\mathbf{F}^D_{\text{viscous}}[\phi_k]$$

$$= \int_D |\phi_k - \mathrm{I\!I}|^2\theta_{k-1} + \frac{1}{\delta}|\theta_k - (\det(D\phi_k)^{-1}\theta_{k-1}) \circ \phi_k^{-1}|^2 \det(D\phi_k) \circ \phi_k^{-1}\,\mathrm{d}x + \gamma\mathbf{F}^D_{\text{viscous}}[\phi_k]\,.$$

Hence, the Euler-Lagrange equation for θ_k is

$$0 = |\phi_{k+1} - \mathrm{I\!I}|^2 - \frac{2}{\delta}(\det(D\phi_{k+1})\theta_{k+1} \circ \phi_{k+1} - \theta_k) + \frac{2}{\delta}(\theta_k - (\det(D\phi_k)^{-1}\theta_{k-1}) \circ \phi_k^{-1})\det(D\phi_k) \circ \phi_k^{-1}$$

for all $k = 1, \ldots, K-1$ and a.e. $x \in D$. Now we define the discrete transport path $X(x) = (X_1(x), X_2(x), \ldots, X_{K-1}(x))$ with $X_1(x) = \phi_1(x)$ and $X_k(x) = \phi_k(X_{k-1}(x))$ and the vector

$$\bar{\theta}(x) = (\theta_1(X_1(x)), \theta_2(X_2(x)), \ldots, \theta_{K-1}(X_{K-1}(x)))\,.$$

Then we can write the optimality conditions as

$$\theta_k = \frac{\det(D\phi_{k+1})\theta_{k+1} \circ \phi_{k+1} + \theta_{k-1} \circ \phi_k^{-1} - \frac{\delta}{2}|\phi_{k+1} - \mathrm{I\!I}|^2}{1 + \det(D\phi_k) \circ \phi_k^{-1}}\,.$$

From $X_k \in C^{1,\alpha}$ we deduce that

$$\theta_k \circ X_k = \frac{(\det(D\phi_{k+1}) \circ X_k)(\theta_{k+1} \circ X_{k+1}) + \theta_{k-1} \circ X_{k-1} - \frac{\delta}{2}|X_{k+1} - X_k|^2}{1 + \det(D\phi_k) \circ X_{k-1}} \tag{24}$$

for a.e. $x \in D$ and for all $k = 1, \ldots, K-1$. This can be rewritten as a linear system $\mathbf{A}(x)\bar{\theta}(x) = \mathbf{R}(x)$, where $\mathbf{A}(x)$ is a tridiagonal matrix given by

$$\mathbf{A}(x)_{k,k+1} = -\frac{\det(D\phi_{k+1}) \circ X_k(x)}{1 + \det(D\phi_k) \circ X_{k-1}(x)}, \quad \mathbf{A}(x)_{k,k} = 1, \quad \mathbf{A}(x)_{k,k-1} = -\frac{1}{1 + \det(D\phi_k) \circ X_{k-1}(x)}$$

and $\mathbf{R}(x) = \mathbf{B}(x) + \mathbf{T}(x)$ with $\mathbf{B}(x), \mathbf{T}(x) \in \mathbb{R}^{K-1}$ given by

$$\mathbf{B}(x) = \left(\frac{\theta_A(x)}{1 + \det(D\phi_1)(x)}, 0, \ldots, 0, \frac{\det(D\phi_K) \circ X_{K-1}(x)\theta_B \circ X_K(x)}{1 + \det(D\phi_{K-1}) \circ X_{K-2}(x)}\right)^T$$

$$\mathbf{T}(x)_k = -\frac{\delta}{2}\frac{|X_{k+1}(x) - X_k(x)|^2}{1 + \det(D\phi_k) \circ X_{k-1}(x)} \quad \forall k = 1, \ldots, K-1\,.$$

Now, the unique minimizer $(\theta_1, \ldots, \theta_{K-1}) \subset \mathcal{I}^{K-1}$ satisfies for a.e. $x \in D$ the derived linear system of equations and gives the only solution of this system. Thus $A(x)$ is invertible for a.e. $x \in D$ and by solving the system we can recover the minimizer. In the constraint case $\theta \geq 0$ a.e. the minimization with respect to $(\theta_1, \ldots, \theta_{K-1})$ no longer decomposes into a linear system of equations with unknowns $(\theta_1(X_1(x)), \ldots, \theta_{K-1}(X_{K-1}(x)))$ for a.e. $x \in D$. But the decomposition along the discrete paths $(X_0(x), \ldots, X_K(x))$ is still applicable. Indeed, one observes that for a.e. $x \in D$ the vector $(\theta_1(X_1(x)), \ldots, \theta_{K-1}(X_{K-1}(x)))$ minimizes the quadratic functional

$$Q(\tilde{\theta}_1, \ldots, \tilde{\theta}_{K-1}) = \sum_{k=1}^{K} \det(DX_{k-1})(x) \left(|X_k(x) - X_{k-1}(x)|^2 \tilde{\theta}_{k-1} + \frac{1}{\delta} |\det(D\phi_k)(X_{k-1}(x))\tilde{\theta}_k - \tilde{\theta}_{k-1}|^2 \right)$$

with $\tilde{\theta}_0 = \theta_A(x)$ and $\tilde{\theta}_K = \theta_B(x)$ over all $(\tilde{\theta}_1, \ldots, \tilde{\theta}_{K-1}) \in \mathbb{R}^{K-1}$ subject to the constraint $\tilde{\theta}_k \geq 0$ for all $k = 1, \ldots, K-1$. This is a simple quadratic optimization problem in $\mathbb{R}^{K-1}$ with inequality constraints.

5 Spatial discretization

With respect to the spatial discretization we follow the procedure already proposed in [BER15]. We restrict to two dimensional images ($d = 2$) and consider a regular quadrilateral grid on the two-dimensional image domain $D = [0,1]^2$ consisting of rectangular cells $\{C_m\}_{m \in I_C}$ with I_C being the associated index set. Let $\mathcal{V}_h$ be the space of piecewise bilinear continuous functions and denote by $\{\xi^i\}_{i \in I_N}$ the set of nodal basis functions with I_N being the index set of all grid nodes x_i.

We investigate spatially discrete deformations $\Phi_k : D \to D$ with $\Phi_k \in \mathcal{V}_h^2 = \mathcal{V}_h \times \mathcal{V}_h$ and spatially discrete image maps $\Theta_k : D \to \mathbb{R}$ with $\Theta_k \in \mathcal{V}_h$. Given any finite element function $U \in \mathcal{V}_h$ we denote by $\bar{U} = (U(x_i))_{i \in I_N}$ the corresponding vector of nodal values. Now, we define a fully discrete counterpart $\mathbf{E}_{\delta,\gamma,h}^K$ of the so far solely time discrete path energy $\mathbf{E}_{\delta,\gamma}^K$ defined in (18) as follows

$$\mathbf{E}_{\delta,\gamma,h}^K[(\Theta_0, \ldots, \Theta_K)] = \min_{\substack{\Phi_k \in \mathcal{V}_h^2 \\ \Phi_k|_{\partial D} = \mathbb{1}}} \mathbf{E}_{\delta,\gamma,h}^{K,D}[(\Theta_0, \ldots, \Theta_K), (\Phi_1, \ldots, \Phi_K)]$$

and obtain the resulting fully discrete approximation of the squared Riemannian distance

$$\mathcal{W}_{\delta,\gamma,h}^K[\Theta_A, \Theta_B]^2 = \min_{\substack{\Theta_0, \ldots, \Theta_K \in \mathcal{I} \\ \Theta_0 = \Theta_A, \, \Theta_K = \Theta_B}} \mathbf{E}_{\delta,\gamma,h}^K[\Theta_0, \ldots, \Theta_K].$$

Here, $\mathbf{E}_{\delta,\gamma,h}^{K,D}[(\Theta_0, \ldots, \Theta_K), (\Phi_1, \ldots, \Phi_K)]$ is the discrete counterpart of $\mathbf{E}_{\delta,\gamma}^{K,D}$ in (21) obtained by the evaluation of all the integrals in $\mathbf{E}_{\delta,\gamma}^{K,D}$ using third order Simpson quadrature with 9 quadrature points. Then, the resulting entries of the weighted mass matrix $\mathbf{M}_h[\omega, \Phi, \Psi]_{ij} = (\mathbf{M}_h[\omega, \Phi, \Psi]_{ij})_{i,j \in I_N}$ with weight ω and transformed via deformations Φ, Ψ are given by

$$\mathbf{M}_h[\omega, \Phi, \Psi]_{ij} = \sum_{l \in I_C} \sum_{q=0}^{8} w_q^l \, \omega(x_q^l) \, (\xi^i \circ \Phi)(x_q^l) \, (\xi^j \circ \Psi)(x_q^l).$$

Here, the x_q^l are the quadrature points and the w_q^l are corresponding quadrature weights. In the case $\omega = 1$ we write $\mathbf{M}_h[1, \Phi, \Psi] = \mathbf{M}_h[\Phi, \Psi]$.

To compute a minimizer of the fully discrete energy $\mathbf{E}_{\delta,\gamma,h}^K$ we proceed as in the existence proof of time discrete geodesics in Section 4 and alternate the optimisation of the set of deformations for fixed image intensities and the optimization of the image intensities for fixed deformations. The optimization of deformations decouples in time. To calculate an optimal, discrete matching deformation for two consecutive images we use a conjugate gradient method for the fully discrete energy $\mathbf{E}_{\delta,\gamma,h}^{K,D}$. In practice we use the following hyperelastic energy $W(D\phi) = \frac{\mu}{2} \|D\phi\|_F^2 + \frac{\lambda}{4} (\det D\phi)^2 - (\mu + \frac{\lambda}{2}) \log(\det D\phi) - \mu - \frac{\lambda}{4}$ for $\det(D\phi) > 0$ with fixed $\lambda = 10$ and $\mu = 1$ and differing from the assumptions in Section 3 we skip the higher order term $|D^m\phi|^2$. Indeed, the associated regularization experimentally turned out not to be necessary, possibly due to the regularization by the spatial discretization. For a fixed vector of discrete deformations $\Phi = (\Phi_1, \ldots, \Phi_K)$ the minimization of $\mathbf{E}_{\delta,\gamma,h}^{K,D}$ with respect to $\Theta = (\Theta_1, \ldots, \Theta_{K-1})$ leads, as in the spatially continuous case, to a linear system of equations. Indeed, we obtain as the discrete counterpart of $\int_D |\phi_k - \mathbb{1}|^2 \theta_{k-1} \, dx$

$$\sum_{k=1}^{K} \sum_{l \in I_C} \sum_{q=0}^{8} w_q^l \left(|\Phi_k - \mathbb{1}|^2 \bar{\Theta}_{k-1} \right)(x_q^l) = \sum_{k=1}^{K} \mathbf{M}_h[|\Phi_k - \mathbb{1}|^2, \mathbb{1}, \mathbb{1}] \bar{\Theta}_{k-1} \bar{1} \,,$$

13

Figure 1: Optimal transport via translation for two different pairs of images, each of them with identical mass. Left: discrete geodesic between a pair of scaled bump maps with $K = 4$, $\delta = 10^{-1}$, $\gamma = 10^{-1}$ is shown, right: discrete geodesic between a square and a translated square with $K = 4$, $\delta = 10^{-1}$, $\gamma = 10^{-2}$.

with $\bar{1} = (1, \ldots, 1) \in \mathbb{R}^{I_N}$ and as the discrete counterpart of $\int_D |\det(D\phi_k)\theta_k \circ \phi_k - \theta_{k-1}|^2 \, \mathrm{d}x$

$$\sum_{k=1}^{K} \sum_{l \in I_C} \sum_{q=0}^{8} w_q^l \left(|\det(D\Phi_k)(\Theta_k \circ \Phi_k) - \Theta_{k-1}|^2 \right)(x_q^l)$$

$$= \sum_{k=1}^{K} \left(\mathbf{M}_h[(\det D\Phi_k)^2, \Phi_k, \Phi_k] \bar{\Theta}_k \cdot \bar{\Theta}_k - 2\mathbf{M}_h[\det(D\Phi_k), \Phi_k, \mathbb{1}] \bar{\Theta}_k \cdot \bar{\Theta}_{k-1} + \mathbf{M}_h[\mathbb{1}, \mathbb{1}] \bar{\Theta}_{k-1} \cdot \bar{\Theta}_{k-1} \right).$$

Hence, the resulting discretized part of $\mathbf{E}_{\delta,\gamma,h}^{K,D}$ depending on $\bar{\Theta}_k$ is given by

$$\mathbf{M}_h[|\Phi_{k+1} - \mathbb{1}|^2, \mathbb{1}, \mathbb{1}] \bar{\Theta}_k \bar{1} + \frac{1}{\delta}(\mathbf{M}_h[(\det D\Phi_k)^2, \Phi_k, \Phi_k] + \mathbf{M}_h[\mathbb{1}, \mathbb{1}]) \bar{\Theta}_k \cdot \bar{\Theta}_k$$

$$- \frac{2}{\delta} \left(\mathbf{M}_h[\det(D\Phi_k), \Phi_k, \mathbb{1}] \bar{\Theta}_k \cdot \bar{\Theta}_{k-1} + \mathbf{M}_h[\det(D\Phi_{k+1}), \Phi_{k+1}, \mathbb{1}] \bar{\Theta}_{k+1} \cdot \bar{\Theta}_k \right).$$

In what follows, we restrict to the non constraint case minimizing over intensities, which are not necessarily non negative. In fact, in our numerical experiments for $\Theta_A, \Theta_B \geq 0$ and with all deformations being initialized with the identity we did not observe negative density values in the vectors $\bar{\Theta}_k$ for $k = 1, \ldots, K-1$. The implementation of a constraint, quadratic optimization method is work in progress.

For the variation of $\mathbf{E}_{\delta,\gamma,h}^{K,D}$ with respect to the k-th image $\bar{\Theta}_k$ one obtains

$$\partial_{\bar{\Theta}_k} \mathbf{E}_{\delta,\gamma,h}^{K,D} = \mathbf{M}_h[|\Phi_{k+1} - \mathbb{1}|^2, \mathbb{1}, \mathbb{1}] \bar{1} + \frac{2}{\delta}(\mathbf{M}_h[(\det D\Phi_k)^2, \Phi_k, \Phi_k] + \mathbf{M}_h[\mathbb{1}, \mathbb{1}]) \bar{\Theta}_k$$

$$- \frac{2}{\delta} \left(\mathbf{M}_h[\det D\Phi_k, \Phi_k, \mathbb{1}]^T \bar{\Theta}_{k-1} + \mathbf{M}_h[\det D\Phi_{k+1}, \Phi_{k+1}, \mathbb{1}] \bar{\Theta}_{k+1} \right).$$

As a consequence the necessary condition for $\boldsymbol{\Theta} := (\Theta_1, \ldots, \Theta_{K-1})$ to be a minimizer of $\mathbf{E}_{\delta,\gamma,h}^{K,D}$ is a block tridiagonal system of linear equations $\mathbf{A}[\boldsymbol{\Phi}]\bar{\boldsymbol{\Theta}} = \mathbf{R}[\boldsymbol{\Phi}]$, where $\mathbf{A}[\boldsymbol{\Phi}]$ is formed by $(K-1) \times (K-1)$ matrix blocks $\mathbf{A}_{k,k'} \in \mathbb{R}^{I_N \times I_N}$ with

$$\mathbf{A}_{k,k-1} = -\mathbf{M}_h[\det D\Phi_k, \Phi_k, \mathbb{1}]^T, \quad \mathbf{A}_{k,k} = \mathbf{M}_h[(\det D\Phi_k)^2, \Phi_k, \Phi_k] + \mathbf{M}_h[\mathbb{1}, \mathbb{1}],$$

$$\mathbf{A}_{k,k+1} = -\mathbf{M}_h[\det D\Phi_{k+1}, \Phi_{k+1}, \mathbb{1}]$$

and $\mathbf{R}[\boldsymbol{\Phi}] = \mathbf{B}[\boldsymbol{\Phi}] + \mathbf{T}[\boldsymbol{\Phi}]$ consists of $K-1$ vector blocks $\mathbf{R}_k = \mathbf{B}_k + \mathbf{T}_k \in \mathbb{R}^{I_N}$ with $\mathbf{B}_1 = \mathbf{M}_h[\det D\Phi_1, \Phi_1, \mathbb{1}]^T \bar{\Theta}_0$, $\mathbf{B}_2 = \ldots = \mathbf{B}_{K-2} = 0$, $\mathbf{B}_{K-1} = \mathbf{M}_h[\det D\Phi_K, \Phi_K, \mathbb{1}] \bar{\Theta}_K$, and $\mathbf{T}_k = -\frac{\delta}{2} \mathbf{M}_h[|\Phi_{k+1} - \mathbb{1}|^2, \mathbb{1}, \mathbb{1}]^T \bar{1}$ for all $k = 1, \ldots, K-1$.

The energy $\sum_{l \in I_C} \sum_{q=0}^{8} w_q^l \left(|\Phi_k - \mathbb{1}|^2 \Theta_{k-1} + \frac{1}{\delta} |\det D\Phi_k \Theta_k \circ \Phi_k - \Theta_{k-1}|^2 \right)(x_q^l)$ is convex in Θ_k and strictly convex in Θ_{k-1}. Hence, $\mathbf{E}_{\delta,\gamma,h}^{K,D}$ is strictly convex in $\boldsymbol{\Theta}$ and there is a unique minimizer $\boldsymbol{\Theta} = \boldsymbol{\Theta}[\boldsymbol{\Phi}]$ for fixed $\boldsymbol{\Phi}$. This implies that $\mathbf{A}$ is invertible and therefore the resulting solution $\boldsymbol{\Theta}$ coincides with the unique minimizer of $\mathbf{E}_{\delta,\gamma,h}^{K,D}$. Numerically, the corresponding system of linear equations is solved with a conjugate gradient method with diagonal preconditioning. In addition, as an outer iteration of the numerical energy descent scheme we apply a cascadic approach starting on coarse grids and successively refining the grid.

In what follows, we will discuss numerical results obtained by the proposed scheme. We start with two simple transport examples of image densities with identical mass. In Figure 1 the optimal transport geodesics connecting a bump map $f(x) = \exp\left(\left(1 - \sigma^{-2}|x - x_0|^2\right)^{-1}\right) \chi_{B_\sigma(x_0)}$ with centre $x_0 \in D$ and radius $\sigma > 0$ and its translate as well as a characteristic function of a square and its translate are considered for small δ and γ. Indeed, the computed optimal transport constitutes of a translation. Next, we illustrate the role of the source term allowing for density modulation in
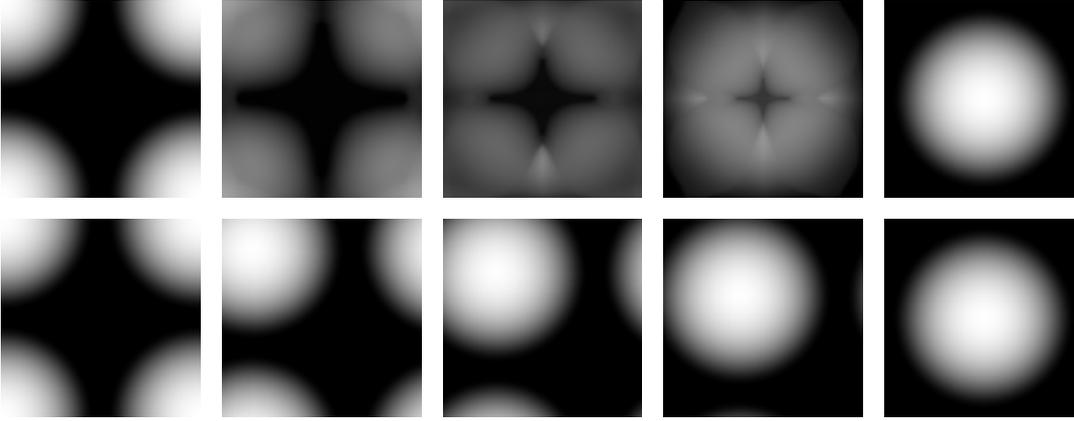
Figure 2: Discrete geodesic are computed for different viscosity parameter γ between two bump maps placed on a square and periodically extended to $\mathbb{R}^2$. Top: $\gamma = 5 \cdot 10^{-4}$, $\delta = 10^{-1}$, bottom: $\gamma = 5$ ($\delta = 10^{-1}$). The images are extracted from a discrete geodesic with $K = 9$. The different contributions to the resulting discrete path energy are: transport cost = 0.0278363, density modulation cost = 0.00107356, dissipation cost = 0.0128489 (top row) and transport cost = 0.155598, density modulation cost = 0.00274782, dissipation cost = 0.00139578 (bottom row).

case of θ_A and θ_B given in Figure 3 as two bump maps of different size at different centre points and in Figure 4 as the characteristic functions of two rectangles of different size still for small γ. In Figure 2 we show the influence of the viscous dissipation. Picking up a test case from [BB00] we consider image intensities on a square periodically extended to $\mathbb{R}^2$ with a bump map once placed in the vertices of the square and once at the centre. We consider periodic boundary conditions both for the image intensities and for the motion field. The Wasserstein geodesic was already computed in [BB00] and we obtain approximately the same result for small γ. Indeed, the bump map at the vertices split up into four pieces, which are then transported separately into the centre. From the perspective of optimal transport this path is energetically preferable due to the shorter transport distance compared to a simple translation from the vertices into the centre. Obviously, this splitting of mass is expensive from the viscous dissipation perspective. Hence, for larger γ we observe the simple translation.

Next, we illustrate the role of the source term allowing for density modulation in case of θ_A and θ_B as in Figure 3 but now with different mass in the two bump maps. Still we impose periodic boundary conditions. For small values of δ we observe a splitting of the bump maps in the corners with the outer one being blended out and the inner one being mainly transported into the middle, whereas for larger values of δ we observe a blending process without significant transport. Furthermore, increasing the viscous dissipation parameter γ leads as in Figure 2 to a translation of the whole bump, while the mass overhead is continuously faded-out. In Figure 4 the input images consists of characteristic functions of two rectangles of different size. Now, we impose natural boundary on ∂D. For small δ and strong penalization of sources the surplus of mass is pushed outwards, whereas for large δ one observes a simple blending and almost no transport.

Furthermore, Figure 5 compares our model with the metamorphosis model on the discrete geodesic between two images consisting of a light and a dark square and the flipped configuration. For very small density modulation parameter ($\delta = 0.01$) we observe a transport of a "light block" from the bottom to the top square, especially mass is approximately preserved. In case of the metamorphism model with purely viscous flow, we see a transport of the lighter square combined with a fading in and out of the darker phase.

As a first imaging application we pick up in Figure 6 an example from [PPO14]. For small δ and small γ we obtain a very similar result. Finally, in Figure 7 the geodesic between two different slices of the same human brain recorder via MRI is shown. The corresponding image intensities are characterized by substantially different masses. In fact, it is the incorporation of both the source term and the viscous dissipation term which enables a reasonable morph between the two slices. Thereby, the source terms allows for local image intensity modulation, whereas the viscous dissipation ensures regularity of the resulting transport path.

There is no guarantee that the alternating algorithm converges. To demonstrate the experimental convergence behaviour we choose the application shown in Fig. 7 and show the evolution of the l^2 norm of the difference between consequitive intensities in Fig. 5.
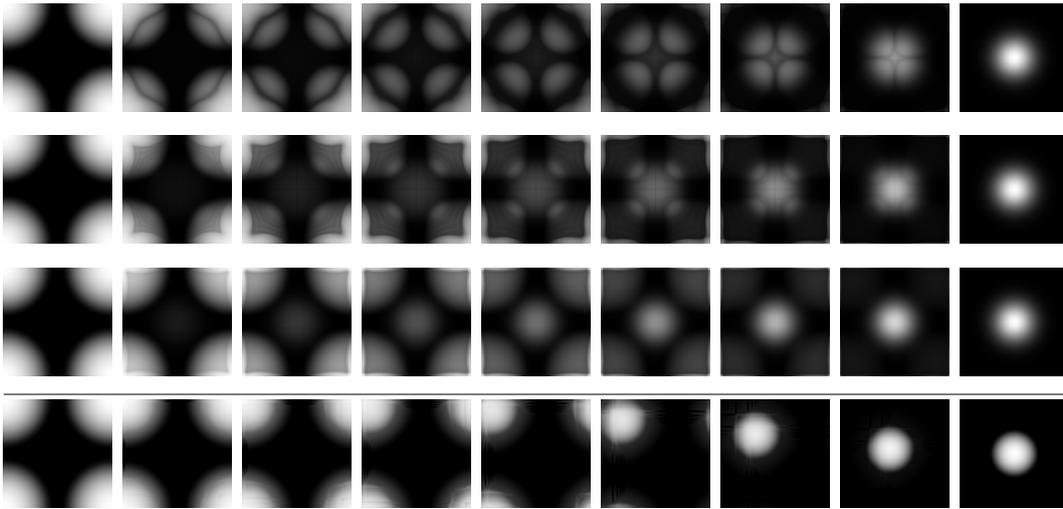
Figure 3: Discrete geodesics ($K = 9$) between two bump maps of different mass for different values of δ (from the first to the third row $\delta = 1, 10, 100$ with $\gamma = 5 \cdot 10^{-4}$). In the fourth row the discrete geodesic for $\gamma = 1$, $\delta = 10^{-1}$ is displayed.
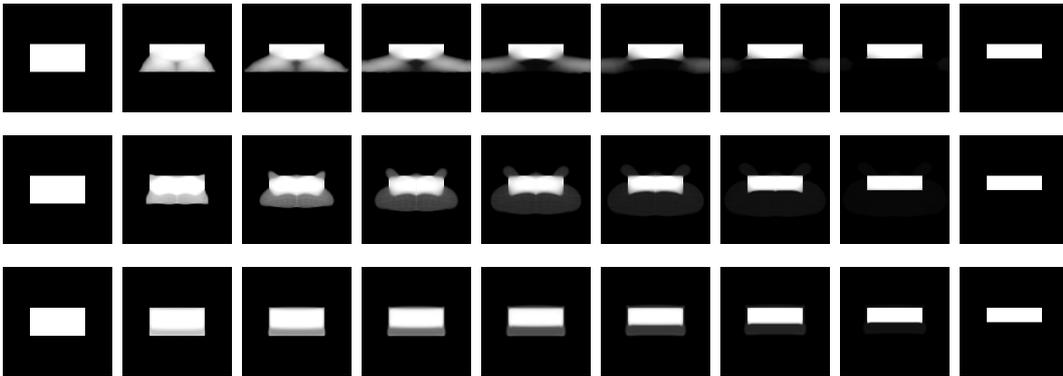

Figure 4: Discrete geodesics between two rectangles of different mass for different values of δ. Top: $\delta = 10^{-2}$, middle: $\delta = 10^{-1}$, bottom: $\delta = 1$ ($K = 9$, $\gamma = 10^{-2}$).
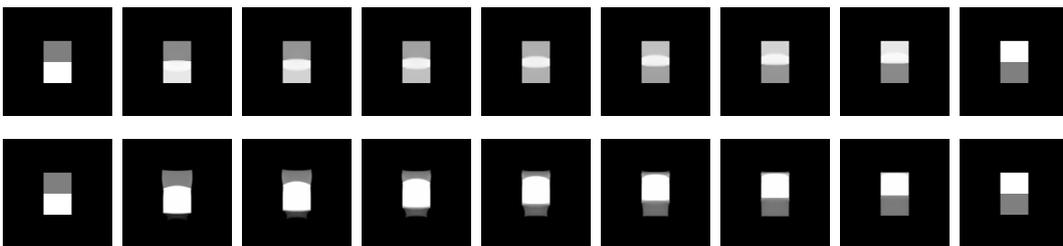
Figure 5: Comparison of the combined model with viscosity parameter $\gamma = 1$ (top) and the metamorphosis model (bottom) for $\delta = 10^{-2}$.
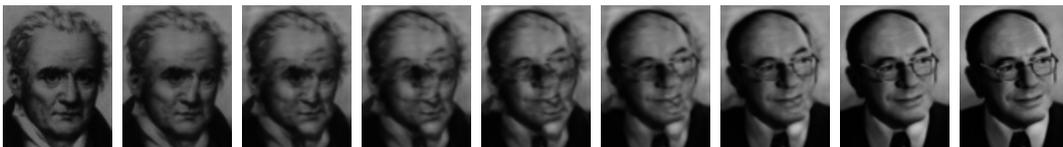
Figure 6: A discrete geodesic between images of Monge and Kantorovich with $\delta = 10^{-2}$, $\gamma = 10^{-2}$ (image provided by G. Peyré).

16
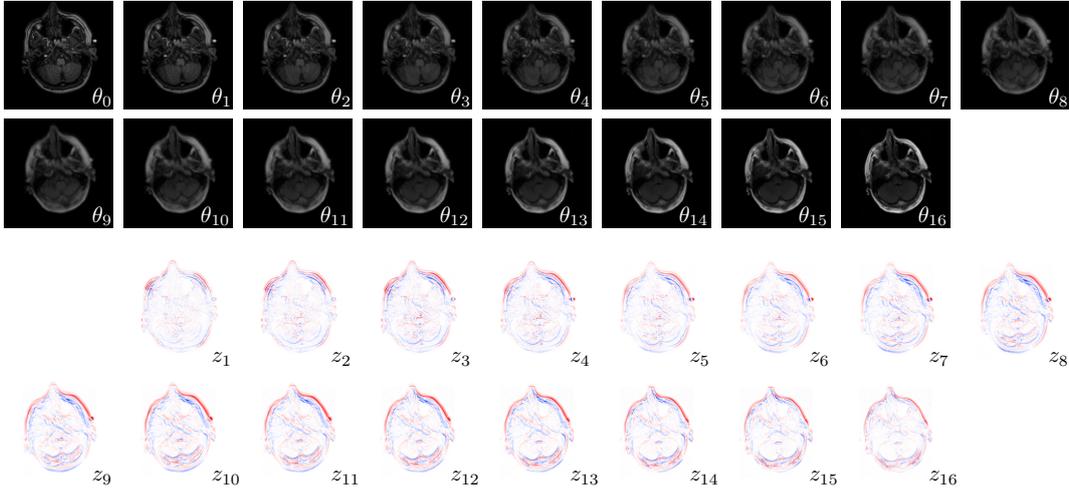
Figure 7: Two slices of the same 3D MRI data set of a human brain are connected with a discrete geodesic (data courtesy of H. Urbach, Neuroradiology, University Hospital Bonn). Top: discrete geodesic with $\delta = 10^{-2}$, $\gamma = 10^{-1}$, bottom: corresponding values of $z_k = \det(D\phi_k)\theta_k \circ \phi_k - \theta_{k-1}$ (blue: positive, red: negative).
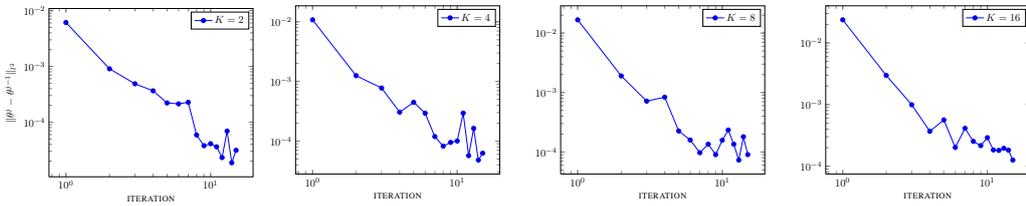

Figure 8: The convergence of the alternating descent method is shown for the application in Fig. 7. For the different levels of the cascadic descent scheme ($K = 2, 4, 8, 16$) the l^2 norm of the difference of consequitive space time densities θ^j is visualized using log-log plots.

6 The Benamou-Brenier discretization for the non viscous model

In this section we numerically compare the proposed approach (11) with the numerical scheme for optimal transport proposed by Benamou and Brenier [BB00], where the mass constraint is relaxed. After the change of variables $(\theta, v) \mapsto (\theta, m = \theta v)$ the minimization problem of the discrete path energy is rewritten as

$$\sup_z \min_\phi \left(F(B\phi) + G(\phi) + \int_0^1 \int_D \phi \cdot z - \frac{1}{\delta}|z|^2 \, dx \, dt \right),$$

where ϕ is a Lagrange multiplier introduced to satisfy the condition on z, F is the indicator function of the convex set $K = \{(a,b) \in \mathbb{R} \times \mathbb{R}^d \; : \; a + \frac{|b|^2}{2} \leq 0\}$, $G(\phi) = \int_D \phi(0,\cdot)\theta_0 - \phi(1,\cdot)\theta_1 \, dx$, and $B : \phi \mapsto (\partial_t \phi, \nabla_x \phi)$. For the outer maximization in z one gets the optimality condition $z = \frac{\delta}{2}\phi$. The augmented Lagrangian is given by

$$L_r[\phi, q, \mu] = F(q) + G(\phi) + \int_0^1 \int_D \phi \cdot z - \frac{1}{\delta}|z|^2 + \mu \cdot (\nabla_{t,x}\phi - q) + \frac{r}{2}|\nabla_{t,x}\phi - q|^2 \, dx \, dt,$$

with variables $q = (a,b)$, $\mu = (\theta, m)$ and Benamou and Brenier propose an alternating gradient descent to compute the saddle point. Using the fact that $z = \frac{\delta}{2}\phi$, one updates z and ϕ simultaneously solving $-r\triangle_{t,x}\phi^n + \frac{\delta}{2}\phi^n = \text{div}_{t,x}(\mu^n - rq^{n-1})$ with Neumann boundary conditions in time, *i.e.* $r\partial_t\phi^n(0,\cdot) = \theta_0 - \theta^n(0,\cdot) + ra^{n-1}(0,\cdot)$, $r\partial_t\phi^n(1,\cdot) = \theta_1 - \theta^n(1,\cdot) + ra^{n-1}(1,\cdot)$. Let us emphasize that in [BB00] the second term on the left hand side which reflects the source term already appeared in the original scheme by Benamou and Brenier as a regularization term.

To study the impact of the parameter δ we pick up the problem already presented in Fig. 2. Now, we choose two input bump maps of different mass. Figure 9 shows discrete geodesics for different δ. For large δ we basically observe pure
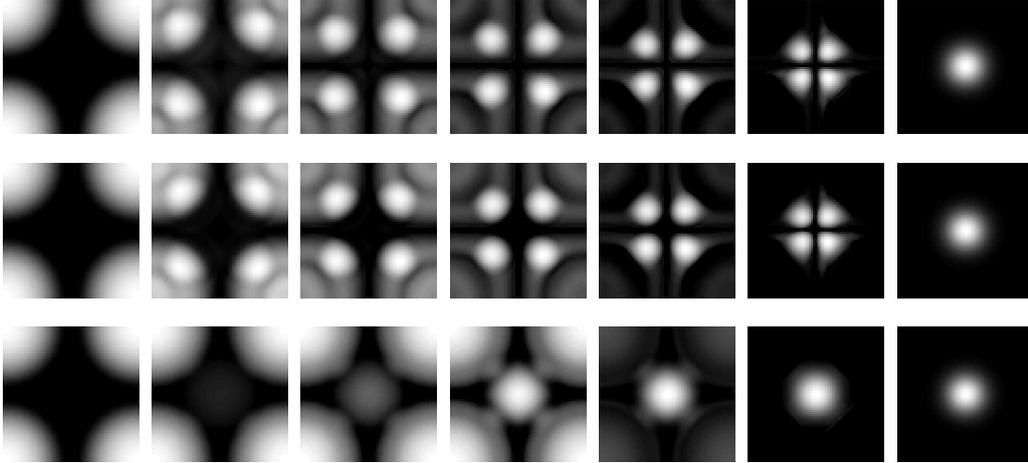

Figure 9: The Benamou-Brenier discretization applied to optimal transport with relaxed mass constraint for input images with bump maps of different mass. The rows show equidistributed time steps for a time step size $\tau = \frac{1}{60}$ and different δ (top: $\delta = 1$, middle: $\delta = 10$, bottom: $\delta = 100$).

blending and almost no transport, whereas for smaller δ mass is first reduced for each bump map leading to a concentration in 4 bumps which are then transported. The differences to Fig. 3 seem to be due to the presence of still some viscous dissipation.

7 Application of the variational time discretization to Riemannian barycentres

As a further application of our time discrete geodesics in the space of images we consider the computation of (weighted) discrete barycentres. We call θ^λ the barycentre of M input images $\theta^1, \ldots, \theta^M$ for given weights $\lambda^1, \ldots, \lambda^M$ with $\lambda^m \geq 0$ and $\sum_{m=1}^M \lambda^m = 1$, if θ^λ minimizes

$$\mathcal{B}_{\delta,\gamma}[\theta] = \sum_{m=1}^M \lambda^m \mathcal{W}_{\delta,\gamma}[\theta, \theta^m]^2$$

Next, replacing the time continuous path energy $\mathcal{W}_{\delta,\gamma}$ by the time discrete energy $\mathcal{W}_{\delta,\gamma}^K$ (19) we ask for a minimizer of the energy

$$\mathcal{B}_{\delta,\gamma}[\theta] = \min_{\substack{(\theta_k^m)_{k=0,\ldots,K} \subset \mathcal{I} \\ (\phi_k^m)_{k=1,\ldots,K} \subset \mathcal{A} \\ \theta = \theta_0^m}} \sum_{m=1}^{M} \lambda^m \mathbf{E}_{\delta,\gamma}^K[\theta_0^m, \ldots, \theta_K^m, \phi_1^m, \ldots, \phi_K^m]$$

over M discrete image paths $(\theta_k^m)_{k=0,\ldots,K}$ $(m = 1, \ldots, M)$ and M discrete families $(\phi_k^m)_{k=1,\ldots,K}$ $(m = 1, \ldots, M)$ with the last image of the m discrete image paths being the mth input image ($\theta_K^m = \theta^m$) and the additional constraint that the set of first images being all equal to θ ($\theta = \theta_0^m$ for all $m = 1, \ldots, M$).

The necessary conditions for the images θ_k^m and the deformations ϕ_k^m ($k = 1, \ldots, M$, $m = 1, \ldots, M$) are identical to those for simple discrete geodesics connecting the corresponding pair of images (θ, θ^m). Solely the condition for the barycentre image itself changes to

$$\theta^\lambda(x) = \max\left(0, \sum_{m=1}^{M} \lambda^m \left(\det(D\phi_1^m)\theta_1^m(\phi_1^m(x)) - \frac{\delta}{2}|\phi_1^m(x) - 1\!\!1|^2\right)\right) .$$

Finally, we take into account the spatial discretization introduced in Section 5 and define Θ^λ as the fully discrete, weighted barycentre of the input images $\Theta^1, \ldots, \Theta^M$, if Θ^λ minimizes the energy

$$\mathbf{B}_{\delta,\gamma,h}^K[\theta] = \sum_{m=1}^{M} \lambda^m \mathcal{W}_{\delta,\gamma,h}^K[\Theta, \Theta^m]^2 \tag{25}$$

$$= \min_{\substack{(\Theta_k^m)_{k=0,\ldots,K} \subset \mathcal{I} \\ (\Phi_k^m)_{k=1,\ldots,K} \subset \mathcal{A} \\ \Theta = \Theta_0^m}} \sum_{m=1}^{M} \lambda^m \mathbf{E}_{\delta,\gamma,h}^K[\Theta_0^m, \ldots, \Theta_K^m, \Phi_1^m, \ldots, \Phi_K^m] .$$

Again for fixed deformations $(\bar{\Phi}_k^m)_{\substack{k=1,\ldots,K, \\ m=1,\ldots,M}}$ and skipping the non negativity constraint for the densities one obtains a system of linear equations to be solved for $(\bar{\Theta}_k^m)_{\substack{k=0,\ldots,K-1, \\ m=1,\ldots,M}}$, with $\bar{\Theta}_0^1 = \ldots = \bar{\Theta}_0^M = \bar{\Theta}^\lambda$. This linear system consists of M copies of the equations for $(\bar{\Theta}_1, \ldots, \bar{\Theta}_{K-1})$ in the system, where we replace $\bar{\Theta}_k$ by $\bar{\Theta}_k^m$ and Φ_k by Φ_k^m, and an additional set of equations for $\bar{\Theta}^\lambda$, *i.e.*

$$\mathbf{M}_h[1\!\!1, 1\!\!1]\bar{\Theta}^\lambda = \sum_{m=1}^{M} \lambda^m \left(\mathbf{M}_h[\det(D\Phi_1^m), \Phi_1^m, 1\!\!1]\bar{\Theta}_1^m - \frac{\delta}{2}\mathbf{M}_h[|\Phi_1^m - 1\!\!1|^2, 1\!\!1, 1\!\!1]\bar{1}\right) .$$

Still, a slightly modified strict convexity argument proves that the energy $\mathbf{E}_{\delta,\gamma,h}^K$ is strictly convex in the images Θ_k^m for $k = 1, \ldots, K$, $m = 1, \ldots, M$ and in the additional image Θ^λ. In particular, there exists a unique solution of the linear system. Let us remark that this is no longer clear if we replace $\mathcal{W}_{\delta,\gamma,h}^K[\Theta^m, \Theta]$ by $\mathcal{W}_{\delta,\gamma,h}^K[\Theta, \Theta^m]$ in the definition of the fully discrete barycenter in (25). In the implementation, we apply an analogous alternating descent scheme as described in Section 5 to compute fully discrete approximations of the weighted Riemannian barycenter. Furthermore, we use a cascadic approach, starting with coarse time discretizations and then successively refining the discretization in time. Figure 10 shows barycenters (with equal weights $\lambda = \frac{1}{M}$) for three different sets of sugar beet slices extracted from noninvasive 3D MRI images at different days after plantation for different viscous dissipation parameters γ. Furthermore, we show the variability of the different contributions to the path energy between the barycenter and the input images for all input sugar beets. Finally, we display in Figure 11 weighted barycenters of three different wood textures with all admissible combinations of $\lambda_m \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$.
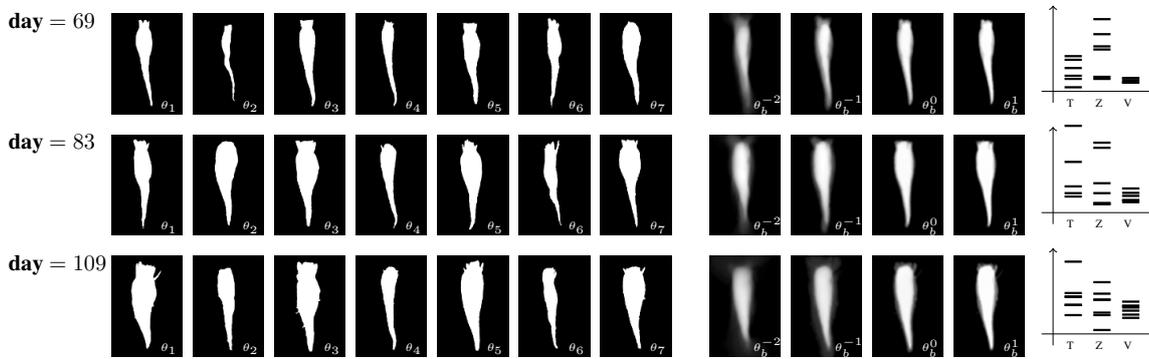
Figure 10: Barycentres of different sets of sugar beet slices (left) for $\delta = 10^{-1}$ and for $\gamma = 10^{-2}$, 10^{-1}, 1, 10 (from left to right with θ_b^j corresponding to $\gamma = 10^j$). On the right the transport cost (T), density modulation cost (Z), and viscous dissipation cost (V) are plotted for all input slices in the case $\gamma = 1$ (cost values for the second beet at $day 69$ are excluded as outliers). (data provided by research network CROP.SENSe.net)
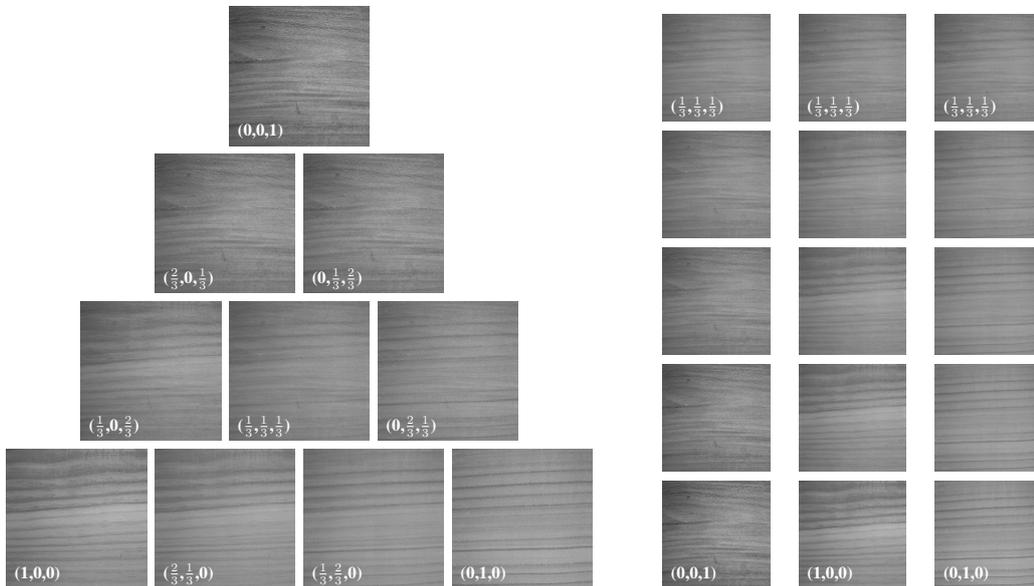

Figure 11: Weighted barycentres of three different wood textures (http://de.wikipedia.org/wiki/Holz) are shown for $\delta = 10^{-1}$, $\gamma = 1$ (Left: barycentric triangle where $(\lambda_0, \lambda_1, \lambda_2)$ are overlaid each texture, right: discrete geodesics for $K = 4$ between the three input textures and the barycentre are depicted.

8 Conclusion and Outlook

In this paper we have developed a combined optimal transport and metamorphosis model and propose an effective time discretization of the path energy in the space of density maps. The method allows us to approximate the original Wasserstein distance and for larger viscosity parameter interesting additional effects can be observed. In particular in applications to images the incorporated source term turns out to be an appropriate way to deal with mass variability. Let us briefly comment on limitations and possible future extensions of the model. So far, in the non-viscous case the source term has to be absolutely continuous with respect to the Lebesque measure (cf. Section 2), because the measure $\mathscr{L}^{\perp}$ in the decomposition of the source measures is not unique. Therefore, a singular part in L^2 would depend on the decomposition. An alternative model with a source term in L^1 including singular parts is work in progress. In addition, in the time discrete model discussed in Section 3 the treatment of the source term in L^2 required special care, since we aim at measuring the change of densities, which is an L^1-concept. Furthermore, for our generalized model including dissipation existence of geodesics in the time continuous case is unclear. In the non-viscous case we made use of a change of variables by considering the momentum instead of the velocity, but for the viscous dissipation term this does not appear to be the appropriate concept. This also renders the verification of Γ-convergence more difficult than in the case of the metamorphosis model in [BER15].

Acknowledgement

References

[AB88]   Luigi Ambrosio and Giuseppe Buttazzo. Weak lower semicontinuous envelope of functionals defined on a space of measures. *Ann. Mat. Pura Appl. (4)*, 150:311–339, 1988.

[AFP00]   Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.

[AGS06]   Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2006.

[AHT03]   Sigurd Angenent, Steven Haker, and Allen Tannenbaum. Minimizing Flows for the Monge–Kantorovich Problem. *SIAM journal on mathematical analysis*, 35(1):61–97, 2003.

[AK98]   V. Arnold and B. Khesin. *Topological methods in hydrodynamics*. Springer, 1998.

[Amb03]   Luigi Ambrosio. Lecture notes on optimal transport problems. Colli, Pierluigi (ed.) et al., Mathematical aspects of evolving interfaces. Lectures given at the C.I.M.-C.I.M.E. joint Euro-summer school, Madeira, Funchal, Portugal, July 3–9, 2000. Berlin: Springer. Lect. Notes Math. 1812, 1-52 (2003)., 2003.

[Arn66]   Vladimir Arnold. Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Annales de l'institut Fourier*, 16:319–361, 1966.

[BB00]   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.

[BCW10]   Martin Burger, José A. Carrillo, and Marie-Therese Wolfram. A mixed finite element method for nonlinear diffusion equations. *Kinet. Relat. Models*, 3(1):59–83, 2010.

[BER15]   B. Berkels, A. Effland, and M. Rumpf. Time Discrete Geodesic Paths in the Space of Images. *ArXiv e-prints*, March 2015.

[BFS12]   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.

[BMTY02] M. F. Beg, M.I. Miller, A. Trouvé, and L. Younes. Computational anatomy: Computing metrics on anatomical shapes. In *Proceedings of 2002 IEEE ISBI*, pages 341–344, 2002.

[BMTY05] M. Faisal Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157, February 2005.

[BS05]   Giuseppe Buttazzo and Filippo Santambrogio. A model for the optimal planning of an urban area. *SIAM J. Math. Anal.*, 37(2):514–530, 2005.

[CEN07]   Tony Chan, Selim Esedoglu, and Kangyu Ni. Histogram based segmentation using Wasserstein distances. In *Scale Space and Variational Methods in Computer Vision*, pages 697–708. Springer, 2007.

[Cia88]   Philippe G. Ciarlet. *Mathematical Elasticity, Volume I: Three-dimensional elasticity*, volume 20 of *Studies in Mathematics and its Applications*. Elsevier, 1988.

[CWVB09] Rick Chartrand, Brendt Wohlberg, Kevin Vixie, and Erik Bollt. A gradient descent solution to the Monge-Kantorovich problem. *Applied Mathematical Sciences*, 3(22):1071–1080, 2009.

[DG06]   Edward J Dean and Roland Glowinski. Numerical methods for fully nonlinear elliptic equations of the Monge–Ampère type. *Computer methods in applied mechanics and engineering*, 195(13):1344–1386, 2006.

[DGM98a] D. Dupuis, U. Grenander, and M.I. Miller. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics*, 56:587–600, 1998.

[DGM98b] Paul Dupuis, Ulf Grenander, and Michael I Miller. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, 56(3):587, 1998.

[DMM10]   Bertram Düring, Daniel Matthes, and Josipa Pina Milišic. A gradient flow scheme for nonlinear fourth order equations. *Discrete Contin. Dyn. Syst. Ser. B*, 14(3):935–959, 2010.

[DNS09]   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.

[Eva99]   Lawrence C. Evans. Partial differential equations and Monge-Kantorovich mass transfer. Bott, Raoul (ed.) et al., Current developments in mathematics, 1997. Papers from the conference held in Cambridge, MA, USA, 1997. Boston, MA: International Press. 65-126 (1999)., 1999.

[FLPJ04] P.T. Fletcher, Conglin Lu, S.M. Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *Medical Imaging, IEEE Transactions on*, 23(8):995–1005, 2004.

[FPR+13] Sira Ferradans, Nicolas Papadakis, Julien Rabin, Gabriel Peyré, and Jean-François Aujol. Regularized discrete optimal transport. In *Scale Space and Variational Methods in Computer Vision*, pages 428–439. Springer, 2013.

[GD04]   Kristen Grauman and Trevor Darrell. Fast contour matching using approximate earth mover's distance. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–220. IEEE, 2004.

[Gre81]   U Grenander. Lectures in pattern theory volume. 1981.

[HRT10]   Eldad Haber, Tauseef Rehman, and Allen Tannenbaum. An efficient numerical method for the solution of the L2 optimal mass transfer problem. *SIAM Journal on Scientific Computing*, 32(1):197–211, 2010.

[HZTA04] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004.

[JM00]   S. C. Joshi and M. I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, 2000.

[Kan42]   Leonid Vital'evich Kantorovitch. On the translocation of masses. *Dokl. Akad. Nauk. USSR*, 37(7-8):227–229, 1942.

[Kan48]   Leonid Vital'evich Kantorovich. On a problem of monge. *Uspekhi Mat. Nauk*, 3:225–226, 1948.

[KMP07]   M. Kilian, N. J. Mitra, and H. Pottmann. Geometric modeling in shape space. In *ACM Transactions on Graphics*, volume 26, pages 1–8, 2007.

[LD11]   Yaron Lipman and Ingrid Daubechies. Conformal Wasserstein distances: Comparing surfaces in polynomial time. *Advances in Mathematics*, 227(3):1047–1077, 2011.

[LO07]   Haibin Ling and Kazunori Okada. An efficient earth mover's distance algorithm for robust histogram comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(5):840–853, 2007.

[LR05]   Grégoire Loeper and Francesca Rapetti. Numerical solution of the Monge–Ampère equation by a Newton's algorithm. *Comptes Rendus Mathematique*, 340(4):319–324, 2005.

[Mem07]   Facundo Memoli. On the use of Gromov-Hausdorff distances for shape comparison. In *Eurographics symposium on point-based graphics*, pages 81–90. The Eurographics Association, 2007.

[Mém11]   Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011.

[Mon81]   Gaspard Monge. *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale, 1781.

[Mos65]   Jürgen Moser. On the volume elements on a manifold. *Transactions of the American Mathematical Society*, 120(2):286–294, 1965.

[MTY02]   M.I. Miller, A. Trouvé, and L. Younes. On the metrics and Euler-Lagrange equations of computational anatomy. *Annual Review of Biomedical Enginieering*, 4:375–405, 2002.

[MY01]   M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: a general framework. *International Journal of Computer Vision*, 41(1–2):61–84, 2001.

[NBCE09]   Kangyu Ni, Xavier Bresson, Tony Chan, and Selim Esedoglu. Local histogram based segmentation using the Wasserstein distance. *International Journal of Computer Vision*, 84(1):97–111, 2009.

[Nv91]   J. Nečas and M. Šilhavý. Multipolar viscous fluids. *Quarterly of Applied Mathematics*, 49(2):247–265, 1991.

[Obe08]   Adam M Oberman. Wide stencil finite difference schemes for the elliptic Monge-Ampere equation and functions of the eigenvalues of the Hessian. *Discrete Contin. Dyn. Syst. Ser. B*, 10(1):221–238, 2008.

[OJBS12]   Laurent Oudre, Jérémie Jakubowicz, Pascal Bianchi, and Chantal Simon. Classification of periodic activities using the Wasserstein distance. *Biomedical Engineering, IEEE Transactions on*, 59(6):1610–1619, 2012.

[PFR12]   Gabriel Peyré, Jalal Fadili, and Julien Rabin. Wasserstein active contours. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2541–2544. IEEE, 2012.

[PPKC10]   Gabriel Peyré, Mickael Péchaud, Renaud Keriven, and Laurent D Cohen. Geodesic methods in computer vision and graphics. *Foundations and Trends® in Computer Graphics and Vision*, 5(3–4):197–397, 2010.

[PPO14]   Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[RP11]   Julien Rabin and Gabriel Peyré. Wasserstein regularization of imaging problem. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1541–1544. IEEE, 2011.

[RPC10]   Julien Rabin, Gabriel Peyré, and Laurent D Cohen. Geodesic shape retrieval via optimal mass transport. In *Computer Vision–ECCV 2010*, pages 771–784. Springer, 2010.

[RPDB12]   Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2012.

[RTG00]   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.

[RW14] Martin Rumpf and Benedikt Wirth. Variational time discretization of geodesic calculus. *IMA Journal of Numerical Analysis*, 2014. (to appear).

[SAK10] Louis-Philippe Saumier, Martial Agueh, and Boualem Khouider. An efficient numerical algorithm for the L2 optimal transport problem with applications to image processing. *arXiv preprint arXiv:1009.6039*, 2010.

[SS13a] Bernhard Schmitzer and Christoph Schnörr. A Hierarchical Approach to Optimal Transport. In *Scale Space and Variational Methods in Computer Vision*, pages 452–464. Springer, 2013.

[SS13b] Bernhard Schmitzer and Christoph Schnörr. Modelling convex shape priors and matching based on the Gromov-Wasserstein distance. *Journal of mathematical imaging and vision*, 46(1):143–159, 2013.

[SS13c] Bernhard Schmitzer and Christoph Schnörr. Object segmentation by shape matching with Wasserstein modes. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 123–136. Springer, 2013.

[Str45] J.W. Strutt. *Theory of sound: Vol. 2*. Dover Publications, 1945.

[TY05] Alain Trouvé and Laurent Younes. Metamorphoses through Lie group action. *Foundations of Computational Mathematics*, 5(2):173–198, 2005.

[Vil03] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2008.

[WBRS11] Benedikt Wirth, Leah Bar, Martin Rumpf, and Guillermo Sapiro. A continuum mechanical approach to geodesics in shape space. *International Journal of Computer Vision*, 93(3):293–318, 2011.

[ZHT03] Lei Zhu, Steven Haker, and Allen Tannenbaum. *Area-preserving mappings for the visualization of medical structures*. Springer, 2003.

[ZYHT07] Lei Zhu, Yan Yang, Steven Haker, and Allen Tannenbaum. An image morphing technique based on optimal mass preserving mapping. *IEEE Transactions on Image Processing*, 16(6):1481–1495, 2007.