Diffeomorphic Density Matching by Optimal Information Transport[*]

Martin Bauer[†], Sarang Joshi[‡], and Klas Modin[§]

Abstract. We address the following problem: given two smooth densities on a manifold, find an optimal diffeomorphism that transforms one density into the other. Our framework builds on connections between the Fisher–Rao information metric on the space of probability densities and right-invariant metrics on the infinite-dimensional manifold of diffeomorphisms. This *optimal information transport*, and modifications thereof, allow us to construct numerical algorithms for density matching. The algorithms are inherently more efficient than those based on optimal mass transport or diffeomorphic registration. Our methods have applications in medical image registration, texture mapping, image morphing, nonuniform random sampling, and mesh adaptivity. Some of these applications are illustrated in examples.

Key words. density matching, information geometry, Fisher–Rao metric, optimal transport, image registration, diffeomorphism groups, random sampling

AMS subject classifications. 58E50, 49Q10, 58E10

DOI. 10.1137/151006238

1. Introduction. In this paper we study the problem of finding diffeomorphic (bijective and smooth) transformations between two densities (volume forms) on an n–manifold M equipped with a Riemannian metric $\mathbf{g}$ and volume form vol. This has applications in many image analysis problems and is an extension of the classical image registration problem. Specific applications of density matching include medical image registration [17, 37, 39, 16], texture mapping in computer graphics [10, 41], image morphing techniques [42, 33], random sampling in Bayesian inference [29, 34], and mesh adaptivity in computational methods for PDEs [9]. A more extensive list of applications and algorithms is given in [32].

The difference between classical image registration (cf. [40]) and density matching is the way transformations act. In image registration, transformations act on positive scalar functions (images) by composition from the right. In density matching, transformations act by

pullback of volume forms: if the density is represented by a function $I \colon M \to \mathbb{R}^+$, the action is given by

$$(1a) \qquad\qquad\qquad\qquad (\varphi, I) \mapsto |D\varphi| I \circ \varphi,$$

where $\varphi \colon M \to M$ is the transformation and $|D\varphi|$ is its Jacobian determinant.

When studying geometric aspects of density matching, it is convenient to use the framework of exterior calculus of differential forms. A density is then thought of as a volume form $\mu = I \mathrm{vol}$, and the action (1a) is given by pullback

$$(1b) \qquad\qquad\qquad\qquad (\varphi, \mu) \mapsto \varphi^* \mu = (|D\varphi| I \circ \varphi)\mathrm{vol}.$$

Notice that the function I is the Radon–Nikodym derivative of μ with respect to vol. For convenience, we use both the function and the exterior calculus point of view throughout the paper; the relation between functions and volume forms is always understood to be $\mu = I \mathrm{vol}$. Pullback (1) is a right action. Sometimes, we shall instead use the corresponding left action, given by pushforward

$$\varphi_* \mu := (\varphi^{-1})^* \mu.$$

Let $\mathrm{Diff}(M)$ and $\mathrm{Dens}(M)$, respectively, denote diffeomorphisms and normalized, smooth densities on M. Both $\mathrm{Diff}(M)$ and $\mathrm{Dens}(M)$ are infinite-dimensional manifolds (see section 2 for details). Let G denote a Riemannian metric on $\mathrm{Diff}(M)$ with corresponding distance square function

$$d^2(\varphi, \psi) := \inf_{\gamma(0)=\varphi, \gamma(1)=\psi} \int_0^1 G_{\gamma(t)}\big(\dot{\gamma}(t), \dot{\gamma}(t)\big)\, \mathrm{d}t.$$

Likewise, let $\bar{G}$ denote a Riemannian metric on $\mathrm{Dens}(M)$ with distance square function $\bar{d}^2$. We are interested in special cases of the following two, generally formulated density matching problems.

Problem 1 (exact density matching). *Given $\mu_0, \mu_1 \in \mathrm{Dens}(M)$, find $\varphi \in \mathrm{Diff}(M)$ minimizing $d^2(\mathrm{id}, \varphi)$ under the constraint*

$$\mu_1 = \varphi_* \mu_0.$$

Equivalently, using intensity functions I_0 and I_1, the constraint is

$$I_1 = |D\varphi^{-1}|(I_0 \circ \varphi^{-1}).$$

Problem 2 (inexact density matching). *Given $\mu_0, \mu_1 \in \mathrm{Dens}(M)$, find $\varphi \in \mathrm{Diff}(M)$ minimizing*

$$(2) \qquad\qquad E(\varphi) := \sigma d^2(\mathrm{id}, \varphi) + \bar{d}^2(\varphi_* \mu_0, \mu_1), \quad \sigma > 0\,.$$

The first term in (2) is a regularity measure, the second term is a similarity measure. The parameter σ is balancing the two criteria.

There is no intrinsic choice of G and $\bar{G}$; they are free to be specified and evaluated in the specific application. The following choices are, however, typically considered.

(i) For Problem 1, the standard choice is distance-squared *optimal mass transport* (OMT), corresponding to the noninvariant L^2 metric

$$G_\varphi(U, V) = \int_M \mathsf{g}(U, V)\mu_0, \quad U, V \in T_\varphi \text{Diff}(M),$$
$$= \int_M \mathsf{g}(u, v)\varphi_*\mu_0, \quad u := U \circ \varphi^{-1}, v := V \circ \varphi^{-1}.$$

This choice of metric induces the Wasserstein distance on $\text{Dens}(M)$ [31, 38].

(ii) For Problem 2, a common choice [35, 20] is the right-invariant H_α^k metric

$$G_\varphi(U, V) = \int_M \mathsf{g}\left((1 - \alpha\Delta)^k u, v\right)\text{vol},$$

where Δ is the Laplace–de Rham operator lifted to vector fields (see subsection 2.1 for details), and, as a similarity measure, the L^2 norm

$$\bar{d}^2(\mu_0, \mu_1) = \|\mu_0 - \mu_1\|_{L^2}^2 := \int_M |I_0 - I_1|^2 \text{vol}.$$

This setting is similar to *large deformation diffeomorphic metric matching* (LDDMM) [21, 26, 35], but with the density action (1) instead of the composition action.

Both choices (i) and (ii) are computationally challenging, as they require the numerical solution of nonlinear partial differential equations (the Monge–Ampere and EPDiff equations, respectively). See [32] and [40] for efficient and stable implementations.

In this paper we consider metrics for Problems 1 and 2 that reduce the computational challenge to solving Poisson equations, allowing significantly faster, semiexplicit algorithms. Our approach is based on connections between information geometry and geodesic equations on diffeomorphism groups (cf. *topological hydrodynamics* [2, 23]).

For the similarity measure on the space of densitites $\text{Dens}(M)$, we consider the infinite-dimensional version of the *Fisher–Rao metric* (also called *Fisher's information metric*), predominant in information geometry. For a pair of tangent vectors $\alpha = a \,\text{vol}$ and $\beta = b \,\text{vol}$ at the base point $\mu = I\text{vol}$, it is given by

$$(3) \qquad \bar{G}_\mu^F(\alpha, \beta) = \frac{1}{4}\int_M \frac{\alpha}{\mu}\frac{\beta}{\mu}\mu = \frac{1}{4}\int_M \frac{ab}{I}\text{vol}.$$

It can be interpreted as the Hessian of relative entropy, or information divergence. This metric is a canonical metric on the space of densities, as it can be defined without using any additional structure of the base manifold M.

On the group of diffeomorphisms we focus on the *information metric* introduced in [27]:

$$(4) \qquad G_\varphi^I(U, V) = \int_M \mathsf{g}(-\Delta u, v)\text{vol} + \lambda \sum_{i=1}^k \int_M \mathsf{g}(u, \xi_i)\text{vol} \int_M \mathsf{g}(v, \xi_i)\text{vol},$$

where $\lambda > 0$, Δ is the Laplace–de Rham operator lifted to vector fields, and $\xi_1, \ldots, \xi_k$ is an orthonormal basis of the harmonic vector fields on M. Because of the Hodge decomposition

theorem, G^I is independent of the choice of orthonormal basis $\xi_1, \ldots, \xi_k$ for the harmonic vector fields. Note that the information metric depends on the finite-dimensional background metric on the manifold M.

What makes the information metric special? Building on work by Khesin et al. [22], Modin [27] showed that G^I induces the Fisher–Rao metric on the quotient space $\text{Diff}_{\text{vol}}(M) \backslash \text{Diff}(M)$ of right cosets, identifiable with $\text{Dens}(M)$. Here, $\text{Diff}_{\text{vol}}(M) \subset \text{Diff}(M)$ denotes volume-preserving diffeomorphisms with respect to vol. Generically, all right-invariant metrics descend to left cosets. The information metric, however, not only descends to left cosets, but also to right cosets. The right-invariance is thereby not exhausted when taking the quotient to right cosets, which implies that the induced metric on the right cosets remains right invariant. It is this property that makes the geometry associated with the information metric remarkable, and more intricate than for other right-invariant metrics.

The geometric setting studied previously in [22, 27] has a limitation in that the background metric on the manifold has to be compatible with the densities being matched. For many applications in image analysis, this limitation is severe as the background metric reflects properties of the ambient space and cannot be chosen freely. This is the motivation and starting point for the theory described in section 5. The aim is to construct a geometric framework, that still allows for efficient numerical computations, while, at the same time, allowing the freedom to prescribe a background metric that is independent of the densities being matched. Towards this goal we extend the theory of [22, 27] to the case of a noncompatible background metric. To retain efficiency of the algorithms we slightly modify the optimality condition. The computational complexity of the resulting algorithms is significantly lower than those based on OMT or LDDMM. Numerical examples are given in section 6.

Contributions of the article. The emphasis of the paper is on new mathematical ideas for density matching rather than highly developed applications. The examples in sections 4 and 6 are therefore kept simple and illustrative. Our first contribution, in section 3, is the reduction of the infinite-dimensional geometry setting in [22, 27] to efficient numerical algorithms. Our second contribution, in section 5, is a new infinite-dimensional geometric framework that overcomes the requirement of a compatible background metric. Furthermore, in section 5 we also reduce this new geometric setting to efficient numerical algorithms.

2. Geometric foundation.

2.1. Notation.
Throughout the paper, the word "metric" always means "Riemannian metric" and "distance" always means "Riemannian distance."

Let M be an n–dimensional orientable manifold with metric $\mathbf{g}$. Oriented local coordinates are denoted $x^1, \ldots, x^n$. We refer to $\mathbf{g}$ as the *background metric*, to distinguish it from metrics on infinite-dimensional spaces. The background metric $\mathbf{g}$ induces a volume form on M, denoted $\text{vol}_\mathbf{g}$. The expression for $\text{vol}_\mathbf{g}$ is

$$\text{vol}_\mathbf{g} = \sqrt{|\mathbf{g}|} \; dx^1 \wedge \cdots \wedge dx^n,$$

where $|\mathbf{g}|$ denotes the determinant of the metric tensor. When the background metric $\mathbf{g}$ is clear from the context, we write vol instead of $\text{vol}_\mathbf{g}$. The total volume of M with respect to

vol, for now assumed to be finite, is denoted

$$\operatorname{vol}(M) := \int_M \operatorname{vol}.$$

The space of smooth, real valued functions on M is denoted $C^\infty(M)$. The space of smooth vector fields on M is given by

$$\mathfrak{X}(M) = \left\{ u^i \frac{\partial}{\partial x^i};\ u^i \in C^\infty(M) \right\}.$$

If $u, v \in \mathfrak{X}(M)$, we sometimes use the notation $u \cdot v$ instead of $\mathbf{g}(u, v)$. The space of smooth 1–forms (covectors) and n–forms on M are given, respectively, by

$$\Omega^1(M) = \{\omega_i\, \mathrm{d}x^i;\ \omega_i \in C^\infty(M)\}$$

and

$$\Omega^n(M) = \{f\, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n;\ f \in C^\infty(M)\}.$$

The background metric $\mathbf{g}$ induces the *musical isomorphism* $\flat \colon \mathfrak{X}(M) \to \Omega^1(M)$ by

$$u^\flat := u^i \mathbf{g}_{ij}\, \mathrm{d}x^j.$$

Its inverse is denoted $\sharp$ and given by

$$\omega^\sharp := \omega_i \mathbf{g}^{ij} \frac{\partial}{\partial x^j},$$

where $\mathbf{g}^{ij}$ are the elements of the inverse metric tensor.

An n–form can (noncanonically) be identified with a smooth function on M: if

$$\alpha = f\, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n \in \Omega^n(M)$$

and $\mu = \rho\, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ is a volume form (ρ is strictly positive), then $\frac{\alpha}{\mu}$ is defined by

$$\frac{\alpha}{\mu} := \frac{f}{\rho}.$$

Notice that this notation is used in the definition of the Fisher–Rao metric (3).

The infinitesimal action of a vector field u on a volume form μ is given by the Lie derivative

$$\mathcal{L}_u \mu = \operatorname{div}_\mu(u)\mu\,,$$

where $\operatorname{div}_\mu(u)$ is the divergence of u with respect to μ, given by

$$\operatorname{div}_\mu(u) = \frac{1}{\sqrt{|\mathbf{g}|}} \frac{\partial}{\partial x^i} \sqrt{|\mathbf{g}|} u^i.$$

When $\mu = \operatorname{vol}$, we write $\operatorname{div}(u)$ instead of $\operatorname{div}_{\operatorname{vol}}(u)$.

The gradient of a function f on M with respect to $\mathbf{g}$ is defined by

$$\operatorname{grad}_{\mathbf{g}} f = (\mathrm{d}f)^{\sharp} = \frac{\partial f}{\partial x^i} \mathbf{g}^{ij} \frac{\partial}{\partial x^j},$$

where d is the natural differential. Again, if $\mathbf{g}$ is clear from the context we write $\operatorname{grad} f$.

Recall the Laplace operator Δ, defined by $\Delta f = \operatorname{div} \operatorname{grad} f$. We also use Δ to denote the lifted Laplace–de Rham operator on the space of vector fields. To define the Laplace–de Rham operator we need the *codifferential operator* δ. It is the generalization of the divergence operator and defined as the adjoint of the differential operator, i.e., $\langle \omega, \mathrm{d}v \rangle_{L^2} = \langle \delta\omega, v \rangle_{L^2}$ (assuming M has no boundary) and depends on the background metric $\mathbf{g}$. The Laplace–de Rham operator is then defined as $\Delta u = -(\delta \mathrm{d}u^{\flat} + \mathrm{d}\delta u^{\flat})^{\sharp}$. We sometimes denote the Laplacian by $\Delta_{\mathbf{g}}$ when the dependence on $\mathbf{g}$ needs to be stressed. If $(M, \mathbf{g})$ is flat, then $\Delta_{\mathbf{g}}$ on vector fields is the standard Laplacian on functions applied elementwise. In $\mathbb{R}^3$ the Laplace–de Rham operator on the space of vector fields is defined using the notion of curl (unlike grad and div, the operator curl does not generalize to arbitrary Riemannian manifolds). Indeed, the Laplace–de Rham operator on $\mathbb{R}^3$ is given by

$$\Delta u = \operatorname{curl} \operatorname{curl} u - \operatorname{grad} \operatorname{div} u.$$

The formula is valid also in the case when $\mathbb{R}^3$ is equipped with a non-Euclidean metric (in which case $\operatorname{curl} u = (\star \mathrm{d}u^{\flat})^{\sharp}$, where $\star$ is the Hodge star).

The space of harmonic vector fields is given by

$$\mathfrak{H}(M) = \{\xi \in \mathfrak{X}(M); \Delta_{\mathbf{g}} \xi = 0\}.$$

If M is a closed manifold (compact and without boundary), then $\mathfrak{H}(M)$ is finite dimensional. If $M = \mathbb{T}^n$, then $\mathfrak{H}(\mathbb{T}^n)$ consists of the vector fields generating translations. By mutual orthogonality between the subspaces in the Hodge decomposition, it follows that the information metric G^I in (4) is independent of the choice of orthonormal basis $\xi_1, \ldots, \xi_k$ for $\mathfrak{H}(M)$. For more details on the Hodge decomposition see, e.g., [27, section 1.1] for manifolds without boundary, or [24, section 3] for manifolds with boundary.

Throughout the paper we use both left and right quotient sets: if G is a group and $H \subset G$ a subgroup, then by *right cosets* we mean the quotient set

$$H \backslash G := \{H \cdot g; g \in G\},$$

and by *left cosets* we mean the quotient set

$$G/H := \{g \cdot H; g \in G\}.$$

2.2. Space of densities and the Fisher–Rao metric. The space of densities $\operatorname{Dens}(M)$ consists of smooth volume forms with total volume $\operatorname{vol}(M)$:

$$\operatorname{Dens}(M) = \left\{ \mu \in \Omega^n(M); \int_M \mu = \operatorname{vol}(M), \mu > 0 \right\}.$$

We like to think of $\operatorname{Dens}(M)$ as an infinite-dimensional manifold. To make this rigorous, first observe that if M is compact, then the space of top forms $\Omega^n(M)$ is a Fréchet space with the

topology induced by the Sobolev seminorms (see Hamilton [18, Example 1.1.5] for details). Let $c \in \mathbb{R}$. Then the set $\Omega_c^n(M) = \{\alpha \in \Omega^n(M); \int_M \alpha = c\}$ is a closed, affine subspace of $\Omega^n(M)$, and $\mathrm{Dens}(M)$ is an open subset of $\Omega_{\mathrm{vol}(M)}^n(M)$. Therefore, $\mathrm{Dens}(M)$ is a *Fréchet manifold* (cf. [18, Chapter 4]). The closure of $\mathrm{Dens}(M)$ is a Fréchet manifold with boundary, given by

$$\overline{\mathrm{Dens}}(M) = \{\alpha \in \Omega_{\mathrm{vol}(M)}^n(M); \alpha \geq 0\}.$$

Since $\mathrm{Dens}(M)$ is an open subset of a closed, affine subspace of a vector space, its tangent space at μ is given by

$$T_\mu \mathrm{Dens}(M) = \Omega_0^n(M).$$

Notice that $T_\mu \mathrm{Dens}(M)$ is independent of μ, so the tangent bundle is trivial $T\mathrm{Dens}(M) \simeq \mathrm{Dens}(M) \times \Omega_0^n(M)$.

As an alternative to the Fréchet topology discussed here, one might work with the completion of $\mathrm{Dens}(M)$ in the Sobolev H^s topology for a differentiability index s. This space, denoted $\mathrm{Dens}^s(M)$, then becomes a *Banach manifold* (see [27] for details). The results in this paper are valid in both the Fréchet and the Banach category.

In the case when M is not compact, an infinite-dimensional manifold structure can still be given, as discussed in subsection 2.6.

Recall the Fisher–Rao metric $\bar{G}^F$ on $\mathrm{Dens}(M)$ given by (3). This metric is weak in the Fréchet (or Banach) topology. Nevertheless, its geodesics are well-posed. Indeed, the astonishing property of the Fisher–Rao metric is that its geodesics are explicit. Following [14], we introduce the W–map

$$(5) \qquad\qquad W \colon \mathrm{Dens}(M) \to C^\infty(M), \quad \mu \mapsto \sqrt{\frac{\mu}{\mathrm{vol}}}.$$

The infinite-dimensional sphere $S^\infty(M) = \{f \in C^\infty(M); \int_M f^2 \,\mathrm{vol} = \mathrm{vol}(M)\}$ is a submanifold of $C^\infty(M)$ and the image of W is an open subset of $S^\infty(M)$ [22, Theorem 3.2]. Indeed, if $\mu \in \mathrm{Dens}(M)$, then

$$\mathrm{vol}(M) = \int_M \mu = \int_M \frac{\mu}{\mathrm{vol}}\mathrm{vol} = \int_M W(\mu)^2 \mathrm{vol} =: \|W(\mu)\|^2.$$

Let $\alpha \in T_\mu \mathrm{Dens}(M)$ and let $p := T_\mu W \cdot \alpha = \frac{\alpha}{2\mathrm{vol}}\sqrt{\frac{\mathrm{vol}}{\mu}}$. Then

$$\|p\|^2 = \int_M \frac{1}{4}\left(\frac{\alpha}{\mathrm{vol}}\right)^2 \frac{\mathrm{vol}}{\mu}\mathrm{vol} = \frac{1}{4}\int_M \left(\frac{\alpha}{\mu}\right)^2 \mu = \bar{G}_\mu^F(\alpha, \alpha).$$

We have thus showed that W is an isometry between $\mathrm{Dens}(M)$ with the Fisher–Rao metric and an open subset of $S^\infty(M)$. Since the geodesics of the infinite-dimensional sphere $S^\infty(M)$ are explicitly known, we obtain the geodesics on $\mathrm{Dens}(M)$. Indeed, the Fisher–Rao geodesic between μ_0 and μ_1 is given by

$$(6) \qquad [0,1] \ni t \mapsto \left(\frac{\sin((1-t)\theta)}{\sin\theta}f_0 + \frac{\sin(t\theta)}{\sin\theta}f_1\right)^2 \mathrm{vol}, \quad \theta = \arccos\left(\frac{\langle f_0, f_1\rangle_{L^2}}{\mathrm{vol}(M)}\right),$$

where $f_0 = W(\mu_0)$ and $f_1 = W(\mu_1)$.

A direct consequence of formula (6) is an explicit formula for the induced geodesic distance. Indeed, if $\bar{d}_F$ denotes the distance function of the Fisher–Rao metric, then

$$(7) \qquad \bar{d}_F(\mu_0, \mu_1) = \sqrt{\text{vol}(M)} \arccos \left(\frac{1}{\text{vol}(M)} \int_M \sqrt{\frac{\mu_0}{\text{vol}} \frac{\mu_1}{\text{vol}}} \text{vol} \right).$$

As already mentioned, this formula for $\bar{d}_F(\mu_0, \mu_1)$ is the key ingredient for our density matching algorithms. It is a spherical version of the Hellinger distance [22]. In Appendix B we compare the geodesic distance of the Fisher–Rao metric to other commonly used distance functions on the space of probability measures.

Remark 2.1. Recall that the Fisher–Rao metric $\bar{G}^F$ on $\text{Dens}(M)$ is canonical: it does not depend on the choice of background metric. For the W–map in (5), this implies that vol can be any volume form, as long as μ is absolutely continuous with respect to it. In particular, as in Example 3.7 below, it does not have to be the volume form associated with the background metric.

Remark 2.2. In information geometry [1], a finite-dimensional submanifold $\Gamma \subset \text{Dens}(M)$ is called a *statistical manifold*. The Fisher–Rao metric on $\text{Dens}(M)$ induces a metric on Γ; in local coordinates $\theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$ it is given by

$$\begin{aligned} G_{ij}^{\Gamma}(\theta) &= \frac{1}{4} \int_M \frac{\partial \ln p(x, \theta)}{\partial \theta_i} \frac{\partial \ln p(x, \theta)}{\partial \theta_j} p(x, \theta) \text{vol}(x), \\ &= \frac{1}{4} \text{E} \left[\frac{\partial \ln p(x, \theta)}{\partial \theta_i} \frac{\partial \ln p(x, \theta)}{\partial \theta_j} \right] \\ &= \int_M \frac{\partial \sqrt{p(x, \theta)}}{\partial \theta_i} \frac{\partial \sqrt{p(x, \theta)}}{\partial \theta_j} \text{vol}(x), \end{aligned}$$

where $\theta \mapsto p(\cdot, \theta) \text{vol} \in \text{Dens}(M)$ is the local coordinate chart. The tensor field $G_{ij}^{\Gamma}(\theta)$ is the classical information matrix of Fisher [13].

2.3. Group of diffeomorphisms and the information metric. The set of diffeomorphisms on M is denoted $\text{Diff}(M)$; it consists of smooth bijective mappings $M \to M$ with smooth inverses. This set has a natural group structure, by composition of maps. If M is compact, then $\text{Diff}(M)$ is a *Fréchet Lie group* [18, section I.4.6], i.e., a Fréchet manifold where the group operations are smooth mappings. The Lie algebra of $\text{Diff}(M)$ is given by the space $\mathfrak{X}(M)$ of smooth vector fields (tangential if M has a boundary). There is a natural choice of an L^2 inner product on $\mathfrak{X}(M)$, given by

$$(8) \qquad \langle u, v \rangle_{L^2} := \int_M \mathsf{g}(u, v) \text{vol}.$$

The tangent space $T_\varphi \text{Diff}(M)$ consists of maps $U \colon M \to TM$ with $U \circ \varphi^{-1} \in \mathfrak{X}(M)$.

As with the space of densities, one can also chose to work in the Sobolev completion $\text{Diff}^s(M)$. For large enough s, the set $\text{Diff}^s(M)$ is a Banach manifold. It is, however, not a Banach Lie group, because left composition is not smooth, only continuous—an issue to be carefully addressed when deriving rigorous existence results for geodesics equations on $\text{Diff}^s(M)$ (see, for example, [12, 27]). The case of noncompact M is discussed in subsection 2.6.

Recall that we are interested in the information metric G^I on $\mathrm{Diff}(M)$, defined in (4). Again, this metric is weak in the Fréchet (or Banach) topology. Nevertheless, the geodesics are well-posed: local existence and uniqueness of the geodesic equation on $\mathrm{Diff}^s(M)$ with the Banach topology is given in [27, section 3]; the result is extended to $\mathrm{Diff}(M)$ with the Fréchet topology by standard techniques as in [12].

The metric G^I has the property of right invariance: if $U, V \in T_\varphi \mathrm{Diff}(M)$ then

$$G^I_\varphi(U, V) = G^I_{\varphi \circ \psi}(U \circ \psi, V \circ \psi) \quad \forall\, \psi \in \mathrm{Diff}(M).$$

This property implies that the geodesic equation can be stated in terms of the reduced variable $u = \dot\varphi \circ \varphi^{-1} \in \mathfrak{X}(M)$, the so called *EPDiff equation* [19] given by

$$\frac{\mathrm{d}}{\mathrm{d}t}m + \mathcal{L}_u m + m \operatorname{div} u = 0, \quad m^\sharp = Au,$$

where $A \colon \mathfrak{X}(M) \to \mathfrak{X}(M)$ is the inertia operator associated with the inner product $G^I_{\mathrm{id}}(\cdot, \cdot)$ on $\mathfrak{X}(M)$. Explicitly,

$$(9) \qquad Au = -\Delta u + \lambda \sum_{i=1}^k \xi_i \int_M \mathbf{g}(u, \xi_i)\mathrm{vol},$$

where $\xi_1, \ldots, \xi_k$ is a basis for the space of harmonic vector fields $\mathfrak{H}(M)$, orthogonal with respect to the L^2 inner product (8). Since G^I is a nondegenerate metric, the operator A is invertible. Let us now compute its inverse.

First, it follows from the Hodge decomposition of 1–forms that the space of vector fields admits the L^2-orthogonal decomposition $\mathfrak{X}(M) = \mathfrak{E}(M) \oplus \mathfrak{H}(M)$, where $\mathfrak{E}(M)$ is the image of the Laplace–de Rham operator $\Delta \colon \mathfrak{X}(M) \to \mathfrak{X}(M)$. The inertia operator is diagonal with respect to this decomposition: if $u = w + \xi$ are the components, then

$$A(w + \xi) = \Delta w + \lambda\xi, \quad \Delta w \in \mathfrak{E}(M), \lambda\xi \in \mathfrak{H}(M).$$

Since Δ is an automorphism on $\mathfrak{E}(M)$, the inverse $\Delta^{-1} \colon \mathfrak{E}(M) \to \mathfrak{E}(M)$ is well-defined, so

$$A^{-1}(w + \xi) = \Delta^{-1}w + \frac{1}{\lambda}\xi.$$

To compute the components w and ξ of $u \in \mathfrak{X}(M)$, it suffices to first compute

$$\xi = \sum_{i=1}^k \langle u, \xi_i\rangle_{L^2}\xi_i$$

and then set $w = u - \xi$. We have thus computed the inverse $A^{-1} \colon \mathfrak{X}(M) \to \mathfrak{X}(M)$. In the case $M = \mathbb{T}^n$ one can use FFT-based numerical algorithms for computing $A^{-1}u$, as we do in section 6.

2.4. Moser's principal bundle structure. Recall from section 1 that the diffeomorphism group $\mathrm{Diff}(M)$ acts from the right on the space of densities $\mathrm{Dens}(M)$ via (1). This action is not free: the isotropy subgroup $\mathrm{Diff}_\mu(M)$ of $\mu \in \mathrm{Dens}(M)$ (also called stabilizer of μ) consists

of those diffeomorphisms that are volume preserving with respect to μ, given by

$$\mathrm{Diff}_\mu(M) := \{\varphi \in \mathrm{Diff}(M); \varphi^*\mu = \mu\}.$$

The action is, however, transitive, proved by Moser [30] for $\mathrm{Diff}(M)$ and $\mathrm{Dens}(M)$ and by Ebin and Marsden [12] for $\mathrm{Diff}^s(M)$ and $\mathrm{Dens}^{s-1}(M)$; for any pair of densities ν, μ there exists a diffeomorphism φ such that $\varphi^*\nu = \mu$. If we fix a reference density $\mu \in \mathrm{Dens}(M)$, we can therefore identify $\mathrm{Dens}(M)$ with the quotient space of right cosets

$$[\varphi] := \mathrm{Diff}_\mu(M) \circ \varphi \in \mathrm{Diff}_\mu(M)\backslash \mathrm{Diff}(M).$$

Indeed, if $\psi \in [\varphi]$, then $\psi^*\mu = \varphi^*\mu$, so the map

$$\mathrm{Diff}_\mu(M)\backslash \mathrm{Diff}(M) \ni [\varphi] \longmapsto \varphi^*\mu \in \mathrm{Dens}(M)$$

is well-defined. It is injective, by construction, and surjective, by Moser's transitivity result. The projection map

$$\pi_\mu : \mathrm{Diff}(M) \ni \varphi \longmapsto \varphi^*\mu \in \mathrm{Dens}(M)$$

thereby provides a principal bundle structure

(10)
$$\begin{array}{c} \mathrm{Diff}_\mu(M) \hookrightarrow \mathrm{Diff}(M) \\ \downarrow{\scriptstyle \pi_\mu} \\ \mathrm{Dens}(M)\,. \end{array}$$

That is, the preimage $\pi_\mu^{-1}(\lambda)$ of each $\lambda \in \mathrm{Dens}(M)$ is a *fiber* in $\mathrm{Diff}(M)$, and each fiber is parameterized by the left action of $\mathrm{Diff}_\mu(M)$ on $\mathrm{Diff}(M)$; see Hamilton [18, section III.2.5] for details. When $\mu = \mathrm{vol}$, we write π instead of π_{vol}.

Remark 2.3. In subsection 2.2 we mapped $\mathrm{Dens}(M)$ to a subset of $S^\infty(M)$ via the W-mapping in (5). The reason for this map was the simple interpretation of the Fisher–Rao metric in this representation (as the sphere metric). Through this map, $\mathrm{Diff}(M)$ also acts on $S^\infty(M)$. Indeed, for $f \in S^\infty(M)$ and $\varphi \in \mathrm{Diff}(M)$ the action is given by

$$f \star \varphi := \sqrt{|D\varphi|}(f \circ \varphi)\,.$$

As required, $W(\varphi^*\mu) = W(\mu) \star \varphi$.

2.5. Riemannian submersions and descending metrics. In this section we show how the information metric G^I and the Fisher–Rao metric $\bar{G}^F$ give a Riemannian structure to the principal bundle (10). This Riemannian structure is the key to our density matching algorithms in sections 3 and 5.

Moser's result on transitivity implies that $\pi \colon \mathrm{Diff}(M) \to \mathrm{Dens}(M)$ is a submersion: it is smooth and its derivative is surjective at every point. Following the work in [27, section 4], we now show that π is, in fact, a *Riemannian submersion* with respect to $\bar{G}^F$ and G^I.

Let $\mathcal{V}_\varphi \subset T_\varphi \mathrm{Diff}(M)$ denote the *vertical distribution* given by the tangent spaces of the fibers of the principal bundle structure (10):

$$U \in \mathcal{V}_\varphi \iff T_\varphi \pi \cdot U = 0.$$

The *horizontal distribution* $\mathcal{H}_\varphi \subset T_\varphi \mathrm{Diff}(M)$ with respect to the information metric G^I is given by

$$U \in \mathcal{H}_\varphi \iff G^I_\varphi(U, V) = 0 \quad \forall V \in \mathcal{V}_\varphi.$$

In other words, $\mathcal{H}_\varphi$ is the orthogonal complement of $\mathcal{V}_\varphi$. The metric G^I_φ *descends* to $\bar{G}^F$ through the principle bundle structure (10). That is

$$G^I_\varphi(U, V) = \bar{G}^F_{\pi_\mu(\varphi)}(T_\varphi \pi_\mu \cdot U, T_\varphi \pi_\mu \cdot V) \quad \forall U, V \in \mathcal{H}_\varphi.$$

A remarkable property of descending metrics is that initially horizontal geodesics remain horizontal: if $\varphi(t)$ is a geodesic curve and $\dot{\varphi}(0) \in \mathcal{H}_{\varphi(0)}$, then $\dot{\varphi}(t) \in \mathcal{H}_{\varphi(t)}$ for all t. Furthermore, if $\varphi(t) \in \mathrm{Diff}(M)$ is a horizontal geodesic curve, then $\mu(t) := \pi(\varphi(t))$ is a geodesic curve on $\mathrm{Dens}(M)$.

In summary, we have the following result.

Lemma 2.4 (see [27]). *Under the identification* $\mathrm{Dens}(M) \simeq \mathrm{Diff}_{\mathrm{vol}}(M) \backslash \mathrm{Diff}(M)$, *the information metric* G^I, *given by* (4), *descends to the Fisher–Rao metric* $\bar{G}^F$, *given by* (3), *i.e.,* $\pi \colon (\mathrm{Diff}(M), G^I) \to (\mathrm{Dens}(M), \bar{G}^F)$ *is a Riemannian submersion. The horizontal distribution is right invariant, given by*

$$\mathcal{H}_\varphi = \left\{ U \in T_\varphi \mathrm{Diff}(M); U \circ \varphi^{-1} = \mathrm{grad}(f), f \in C^\infty(M) \right\}.$$

Remark 2.5. In [4] it was shown that the Fisher–Rao metric is the unique metric on $\mathrm{Dens}(M)$ that is invariant under the action of the diffeomorphism group. As a consequence, any right-invariant metric on $\mathrm{Diff}(M)$ that descends to a metric on $\mathrm{Dens}(M)$ through the principal bundle structure (10) descends to the Fisher–Rao metric.

Remark 2.6. The condition for a right-invariant metric to descend to right cosets is transversal to the condition for a subgroup to be totally geodesic. See [28] for details.

2.6. Base manifold with infinite volume. In the setting so far we assumed that M is compact and that $\mathrm{vol}(M)$ is finite. In some applications it is useful to drop this assumption and consider noncompact manifolds with infinite volume, in particular $M = \mathbb{R}^n$. By imposing decay conditions on the set of densities and diffeomorphisms, the previously described theory continues to hold, as now briefly explained. For a detailed exposition in the case of the diffeomorphism group on $\mathbb{R}$, see [3].

Let vol be the volume form of a noncompact Riemannian manifold M. We introduce compactly supported densities, diffeomorphisms, and functions via

$$\mathrm{Dens}_c(M) := \left\{ I\, \mathrm{vol} \in \mathrm{Dens}(M); \{x; I(x) \neq 1\} \text{ has compact closure} \right\},$$
$$\mathrm{Diff}_c(M) := \left\{ \varphi \in \mathrm{Diff}(M); \{\varphi; \varphi(x) \neq x\} \text{ has compact closure} \right\},$$
$$C^\infty_c(M) := \left\{ f \in C^\infty(M); f \text{ has compact support} \right\}.$$

With these decay conditions, the theory described in subsections 2.2–2.5 for the compact case extends to the noncompact, infinite volume case. In particular, Moser's lemma is still valid.

Then the space of compactly supported densities is an open subset of a sphere with infinite radius, thus a flat space. To see this, we slightly modify the W-mapping (5) to

$$\tilde{W} \colon \mathrm{Dens}(M) \to C^\infty_c(M), \qquad \mu \mapsto \sqrt{\frac{\mu}{\mathrm{vol}} - 1}\,.$$

Using this mapping, the formula (6) for geodesics on $\mathrm{Dens}_c(M)$ simplifies to

$$[0,1] \ni t \mapsto \tilde{W}^{-1}\left((1-t)\tilde{W}(\mu_0) + t\tilde{W}(\mu_1)\right)$$

and the induced geodesic distance is given by the Hellinger distance.

3. Matching with compatible background metric. In this section we derive efficient algorithms to solve Problems 1 and 2 with respect to the information metric (4) on $\mathrm{Diff}(M)$ and the Fisher–Rao metric (3) on $\mathrm{Dens}(M)$ and a background metric g on M fulfilling the compatibility constraints $\mathrm{vol}_{\mathsf{g}} = \mu_0$. This property is fulfilled in some applications of density matching, for example, texture mapping, random sampling, and mesh adaptivity.

An integral component of our method is the ability to horizontally lift paths in $\mathrm{Dens}(M)$ to the diffeomorphism group. Indeed, the selection of G^I on $\mathrm{Diff}(M)$ and $\bar{G}^F$ on $\mathrm{Dens}(M)$ fulfils two central properties: (i) Fisher–Rao geodesics on $\mathrm{Dens}(M)$ are explicit [14] and (ii) the metrics (4) and (3) are related via a principal bundle structure [27]. These are the properties that allow us to construct fast, explicit algorithms.

3.1. Horizontal lifting of paths of densities. Given a path of densities $\mu(t) \in \mathrm{Dens}(M)$ we want to find a path of diffeomorphisms $\varphi(t)$ that project onto $\mu(t)$ with respect to Moser's principal bundle (10), i.e.,

$$(11) \qquad \pi_{\mu(0)}(\varphi(t)) = \varphi(t)^*\mu(0) = \mu(t).$$

Such a path is not unique, since we can compose any solution φ from the left with any diffeomorphisms $\psi \in \mathrm{Diff}_{\mu(0)}(M)$. To address the nonuniqueness, we therefore consider paths of diffeomorphisms fulfilling (11) while of minimal length with respect to the information metric G^I, given by (4). Mathematically, this problem is formulated as follows.

Problem 3 (horizontal lifting). *Given a path of densities $\mu \in C^1([0,1], \mathrm{Dens}(M))$, find a path of diffeomorphisms $\varphi \in C^1([0,1], \mathrm{Diff}(M))$ fulfilling*

$$\varphi(0) = \mathrm{id},$$
$$(12) \qquad \varphi(t)^*\mu(0) = \mu(t),$$

while minimizing

$$\int_0^1 G^I_{\varphi(t)}(\dot{\varphi}(t), \dot{\varphi}(t)) \, \mathrm{d}t.$$

In general there is no easy way to solve this problem. If, however, the background metric fulfils the following compatibility condition, then Problem 3 reduces to solving Poisson equations.

Definition 3.1 (compatible background metric). *Let $\mu \in \mathrm{Dens}(M)$. The background metric g on M is called* compatible *with μ if $\mathrm{vol}_{\mathsf{g}} = \mu$.*

The following result explains the advantage of having a background metric compatible with $\mu(0)$.

Lemma 3.2. *Let the background metric g on M be compatible with $\mu(0)$. Then there is a unique path of diffeomorphisms solving Problem 3. This path is horizontal with respect to the information metric (4).*

Proof. To prove this statement we differentiate (12) for $\mu(t)$:

$$\dot{\mu}(t) = \partial_t(\varphi(t)^*\mu(0)) = \varphi(t)^*\mathcal{L}_{v(t)}\mu(0). \tag{13}$$

Here $v(t) \in \mathfrak{X}(M)$ denotes the right trivialized derivative of φ, i.e., $\dot{\varphi} \circ \varphi^{-1} = v$. Using the compatibility condition, we get

$$\dot{\mu}(t) = \varphi(t)^*\mathcal{L}_{v(t)}\mathrm{vol}_{\mathbf{g}} = \mathrm{div}(v(t)) \circ \varphi(t)\,\mu(t). \tag{14}$$

By the Hodge–Helmholz decomposition for vector fields, we can write v as

$$v = \mathrm{grad}\,f + w$$

for some function f and a divergence-free part w (notice that w also contains the harmonic part). Equation (13) only determines the divergence part of the vector field v, which allows us to choose w freely. Since the Hodge–Helmholz decomposition is orthogonal with respect to the information metric G^I (see [27, section 2]), the length of φ is minimal for $w = 0$. This is the horizontality condition, described in subsection 2.5. That the solution is unique follows from uniqueness of the information factorization of diffeomorphisms; see [27, section 5]. ∎

From Lemma 3.2 we obtain an equation for the solution of the lifting problem.

Theorem 3.3. *Under the same conditions as in Lemma 3.2, the unique solution of Problem 3 is obtained by solving the Poisson equation*

$$\begin{aligned} \Delta f(t) &= \frac{\dot{\mu}(t)}{\mu(t)} \circ \varphi(t)^{-1}, \\ v(t) &= \mathrm{grad}(f(t)), \\ \dot{\varphi}(t) &= v(t) \circ \varphi(t), \quad \varphi(0) = \mathrm{id}. \end{aligned} \tag{15}$$

Proof. The horizontal bundle is generated by gradient vector fields $v = \mathrm{grad}\,f$ for some function f. Thus,

$$\mathrm{div}(v(t)) = \mathrm{div}(\mathrm{grad}(f)) = \Delta f.$$

From (14) we then obtain (15). ∎

Remark 3.4. Because the solutions to Problem 3 are horizontal, the part of the metric G^I penalizing the divergence-free part does not affect the solutions. In particular, the parameter λ in (4) does not affect the solution.

Remark 3.5. We can use the equivariance of the Laplacian and the gradient to rewrite the differential equation (15) as

$$\begin{aligned} \Delta_{\varphi(t)^*g}(f(t) \circ \varphi(t)) &= \frac{\dot{\mu}(t)}{\mu(t)}, \\ \varphi(t)^*v(t) &= \mathrm{grad}_{\varphi^*g}(f(t) \circ \varphi(t)), \\ \dot{\varphi}(t) &= v(t) \circ \varphi(t), \quad \varphi(0) = \mathrm{id}. \end{aligned}$$

If we introduce $h(t) := f(t) \circ \varphi(t), w(t) := \varphi(t)^*v(t)$, and $g(t) := \varphi(t)^*g$ the above equations

become

$$\Delta_{g(t)}(h(t)) = \frac{\dot{\mu}(t)}{\mu(t)},$$
$$w(t)) = \operatorname{grad}_{g(t)}(h(t)),$$
$$\dot{\varphi}(t) = T_{\operatorname{id}}\varphi^{-1} \cdot w(t), \quad \varphi(0) = \operatorname{id}.$$

The main difference with (15) is the time-dependence of the Laplacian in Poisson's equation.

A numerical algorithm for solving Problem 3 is now given as follows.

ALGORITHM 1. *Assume we have a numerical way to represent functions, vector fields, and diffeomorphisms on M, and numerical methods for* (i) *composing functions and vector fields with diffeomorphisms,* (ii) *computing the gradient of functions, and* (iii) *computing the inverse of the Laplace operator Δ. Given a C^1-path of densities $[0,1] \ni t \mapsto \mu(t)$ with $\mu(0) = \operatorname{vol}$, a numerical algorithm for computing discrete lifted paths $\{\varphi_k\}_{k=0}^N$ and $\{\varphi_k^{-1}\}_{k=0}^N$ is given as follows:*

1. *Choose a step size $\varepsilon = 1/N$ for some positive integer N. Initialize $t_0 = 0$, $\varphi_0 = \operatorname{id}$, and $\varphi_0^{-1} = \operatorname{id}$. Set $k \leftarrow 0$.*
2. *Compute $I_k = \frac{\dot{\mu}(t_k)}{\mu(t_k)} \circ \varphi_k^{-1}$ and solve the Poisson equation*

$$\Delta f_k = I_k.$$

3. *Compute the gradient vector field $v_k = \operatorname{grad} f_k$.*
4. *Construct approximations ψ_k to $\exp(\varepsilon v_k)$ and ψ_k^{-1} to $\exp(-\varepsilon v_k)$, for example,*

$$\psi_k = \operatorname{id} + \varepsilon v_k, \quad \psi_k^{-1} = \operatorname{id} - \varepsilon v_k.$$

5. *Update the diffeomorphisms*

$$\varphi_{k+1} = \psi_k \circ \varphi_k, \quad \varphi_{k+1}^{-1} = \varphi_k^{-1} \circ \psi_k^{-1}.$$

6. *Set $k \leftarrow k + 1$ and continue from step 2 unless $k = N$.*

3.2. Exact compatible density matching (optimal information transport (OIT)). The special case of Problem 1 with G^I and $\bar{G}^F$ for infinite-dimensional metrics and a compatible background metric gives OIT.

Problem 4 (OIT). *Given $\mu_1 \in \operatorname{Dens}(M)$, find $\varphi \in \operatorname{Diff}(M)$ minimizing $d_I(\operatorname{id}, \varphi)$ under the constraint*

(16) $$\mu_1 = \varphi_* \operatorname{vol}.$$

Equivalently, using the density function I_1 for μ_1, the constraint is

$$I_1 = |D\varphi^{-1}|.$$

To better conform to the horizontal lifting setup in subsection 3.1, which uses the pullback rather than pushforward action, we notice that if φ is a solution to Problem 4, then φ^{-1} is a solution to the same problem but with pullback in (16) instead of pushforward. This follows from right invariance of G^I, as $d_I(\operatorname{id}, \varphi) = d_I(\operatorname{id}, \varphi^{-1})$.

The following result is a direct consequence of the information factorization of diffeomorphisms [27, Theorem 5.6].

Theorem 3.6. *Problem* 4 *has a unique solution. Its inverse is given by the endpoint of the solution to the lifting equations* (15) *for the path* $\mu(t)$ *given by the Fisher–Rao geodesic* (6) *with* $\mu_0 =$ vol.

It follows that a numerical algorithm for Problem 4 is given by Algorithm 1 with $\mu(t)$ as in Theorem 3.6. This algorithm is demonstrated in section 4 below. Before that, we solve the lifting equations explicitly in the one-dimensional case.

Example 3.7. We want to explicitly solve Problem 4 in dimension one. Let $\mu_0 = I_0\,\mathrm{d}x$, $\mu_1 = I_1\,\mathrm{d}x$ be two arbitrary densities on $M = S^1 \simeq \mathbb{R}/\mathbb{Z}$. Since M is one dimensional we could solve this problem directly, using that, up to translations, the diffeomorphism φ is determined by the matching constraint only. We shall, however, refrain from using this fact and instead solve the lifting equations (15).

The standard metric on S^1 is not compatible with μ_0 unless $f_0 \equiv 1$. Nevertheless, it is straightforward to construct a compatible background metric: choose $\mathbf{g} = I_0^2\,\mathrm{d}x \otimes \mathrm{d}x$. Then $\mathrm{vol}_{\mathbf{g}} = \mu_0$ as required. In contrast to the higher-dimensional case, this choice of a compatible background metric is uniquely determined by I_0; two different ways to construct compatible metrics in the higher-dimensional case are described in Appendix A.

Using Theorem 3.6 we are now able to obtain the solution. To simplify the notation, let $f_0 = \sqrt{I_0}$ and $f_1 = \sqrt{I_1}$, corresponding to the W–map $W(\mu) = \sqrt{\mu/\mathrm{d}x}$. First we recall the formula (6) for the Fisher–Rao geodesics on $\mathrm{Dens}(S^1)$, i.e.,

$$\mu(t) = \left(\frac{\sin\left((1-t)\theta\right)}{\sin\theta} f_0 + \frac{\sin\left(t\theta\right)}{\sin\theta} f_1 \right)^2 \mathrm{d}x\,,$$

where the angle θ is given by

$$\theta = \arccos\left(\frac{1}{2\pi} \int_0^{2\pi} f_0 f_1\,\mathrm{d}x \right).$$

To calculate the lifting equations we need a formula for the gradient and Laplacian of the metric $\mathbf{g}$. For any function f we have

$$\mathrm{grad}_{\mathbf{g}}(f) = \frac{1}{I_0}\partial_x f, \qquad \Delta_g f = \frac{1}{f_0^2}\partial_x(f_0\partial_x f).$$

These formulas are derived in a more general setting in subsection A.1. To simplify the notation we let

$$h(t,x) = \frac{\dot{\mu}(t)}{\mu(t)} = \frac{2\theta\left(\cos(\theta - t\theta)f_0(x) - \cos(t\theta)f_1(x)\right)}{f_1(x)\sin(t\theta) + \sin(\theta - t\theta)f_0(x)}.$$

The lifting equations (15) now become

$$\varphi_t \circ \varphi^{-1} = \frac{\partial_x f}{f_0}, \quad \varphi(0) = \mathrm{id},$$

$$\frac{1}{f_0^2}\partial_x(f_0\partial_x f) = h \circ \varphi^{-1}.$$

The solution to this PDE is given by

$$\varphi(t,x) = \psi(t, \psi(0,x)^{-1}) \quad \text{with} \quad \psi(t,x) = \int_0^x \frac{\mu(t)}{\mathrm{d}x}\mathrm{d}y\,,$$

where $\frac{\mu(t)}{\mathrm{d}x}$ is the Radon–Nikodym derivative of $\mu(t)$ with respect to $\mathrm{d}x$. Evaluating at $t = 1$

we obtain

$$\varphi(1, x) = \left(\int_0^x f_1 \, \mathrm{d}y \right) \circ \left(\int_0^x f_0 \, \mathrm{d}y \right)^{-1}.$$

3.3. Inexact compatible density matching. We are now interested in the special case of Problem 2 with G^I and $\bar{G}^F$ for infinite-dimensional metrics and a compatible background metric.

Problem 5 (inexact, compatible density matching). *Given $\mu_1 \in \mathrm{Dens}(M)$, find $\varphi \in \mathrm{Diff}(M)$ minimizing*

$$E(\varphi) = \sigma d_I^2(\mathrm{id}, \varphi) + \bar{d}_F^2(\varphi_* \mathrm{vol}, \mu_1),$$

where $\sigma > 0$ is a fixed parameter.

As in Problem 4, φ is a minimizer of $E(\varphi)$ if and only if $\phi = \varphi^{-1}$ is a minimizer of

(17) $$\tilde{E}(\phi) = \sigma d_I^2(\mathrm{id}, \phi) + \bar{d}_F^2(\phi^* \mathrm{vol}, \mu_1).$$

Our approach for Problem 5 is to minimize (17) and then obtain the solution by taking the inverse. From Lemma 2.4, i.e., that G^I descends to $\bar{G}^F$, it follows that the lifting equations can be used also to obtain the solution to Problem 5.

Theorem 3.8. *Problem 5 has a unique solution, obtained as follows. Let $\varphi(t)$ be the horizontally lifted geodesic such that $\varphi(1)^{-1}$ solves Problem 4 for the same μ_1. Let $s = \frac{1}{1+\sigma}$. Then the solution is given by $\varphi(s)^{-1}$.*

Proof. A minimizer ϕ of (17) must be connected to the identity by a horizontal geodesic (otherwise it would be possible to find another diffeomorphism with a strictly smaller value of $\tilde{E}$, using [27, Theorem 5.6]). Therefore, minimizing $\tilde{E}$ is equivalent to minimizing the functional $\tilde{e}(\mu) := \sigma \bar{d}_F^2(\mathrm{vol}, \mu) + \bar{d}_F^2(\mu, \mu_1)$ on $\mathrm{Dens}(M)$. First, notice that $\tilde{e}$ is convex on $\mathrm{Dens}(M)$, so a unique minimizer exists. Denote it ν. From the spherical geometry of $(\mathrm{Dens}(M), \bar{G}^F)$, explained in subsection 2.3, it is clear that ν must belong to the geodesic curve $\mu(t)$ between vol and μ_1. Without loss of generality, we may assume that the distance between vol and μ_1 is 1. Using arc length parameterization $\mu(s)$, we then get $\tilde{e}(\mu(s)) = \sigma s^2 + (1 - s)^2$. Minimization over s now proves the result. ∎

From Theorem 3.8 it follows that a numerical algorithm for Problem 5 is obtained through Algorithm 1 by solving the lifting equations until reaching $t = 1/(\sigma + 1)$ and then taking the inverse.

4. Examples of matching with compatible metric. In this section we give some examples of matching with a compatible metric, using Algorithm 1.
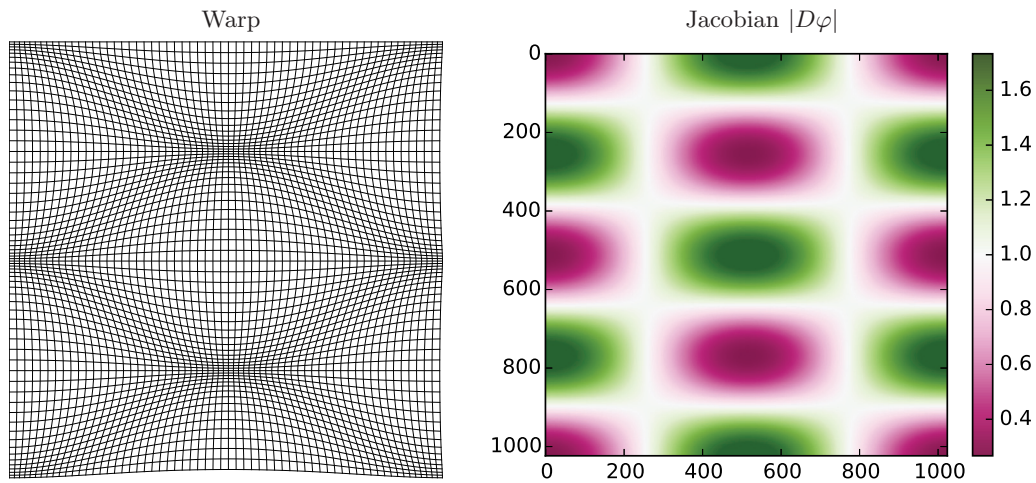
4.1. Random sampling from nonuniform probability distributions. In $\mathbb{R}$, a classical algorithm for generating random samples from an arbitrary probability density function is to use the result in Example 3.7. That is, one uses the cumulative distribution function to transform the standard uniform random variable on the unit interval. Algorithm 1 can analogously be used to transform uniform random samples to samples from an arbitrary probability density on M.

As an example, let $M = \mathbb{T}^2$ (the two-dimensional flat torus). We want to produce random samples from an arbitrary probability distribution on $\mathbb{T}^2$, for example,

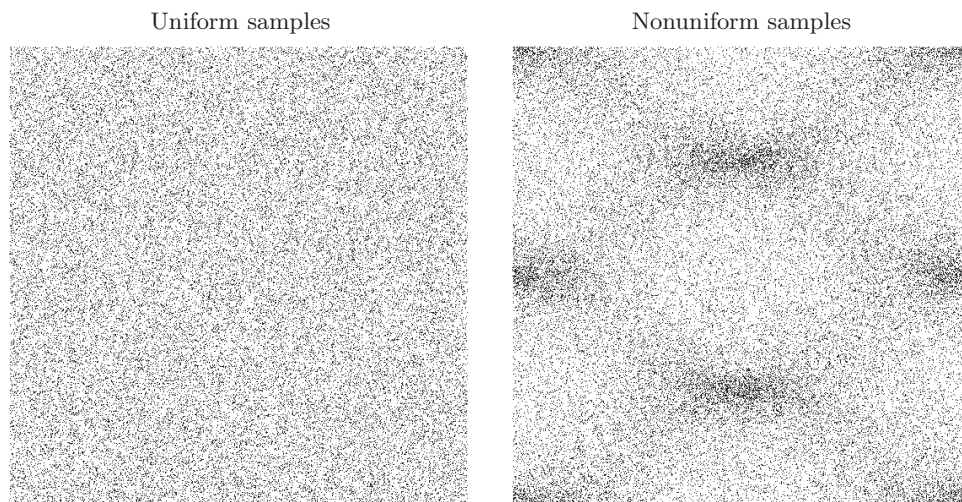$$\mu_1 = \big(1 - 0.8\cos(x)\cos(2y)\big) \underbrace{\mathrm{d}x \wedge \mathrm{d}y}_{\text{vol}},$$

where $x, y \in [-\pi, \pi)$ are coordinates. The approach is to first use the lifting algorithm for Problem 4 to compute $\varphi \in \mathrm{Diff}(\mathbb{T}^2)$ such that $\varphi_*\mathrm{vol} = \mu_1$, then draw random samples (x_i, y_i) from the uniform distribution (using a uniform random number generator), then map these samples into the μ_1-distribution by $(\tilde{x}_i, \tilde{y}_i) = \varphi(x_i, y_i)$.

For a 1024×1024 grid on $\mathbb{T}^2$, with step size $\varepsilon = 0.05$ (a total of 20 steps), we obtain the following warp φ and Jacobian $|D\varphi|$:

Green and pink shades of the Jacobian imply, respectively, expansion and contraction. The optimal information framework assures the warp is matching the two probability densities while locally minimizing metric distortion.
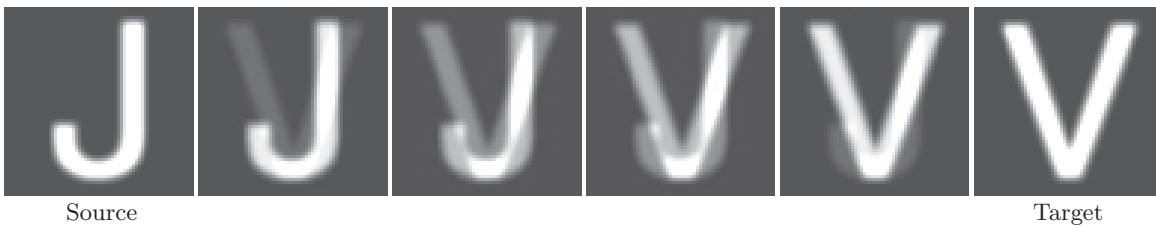
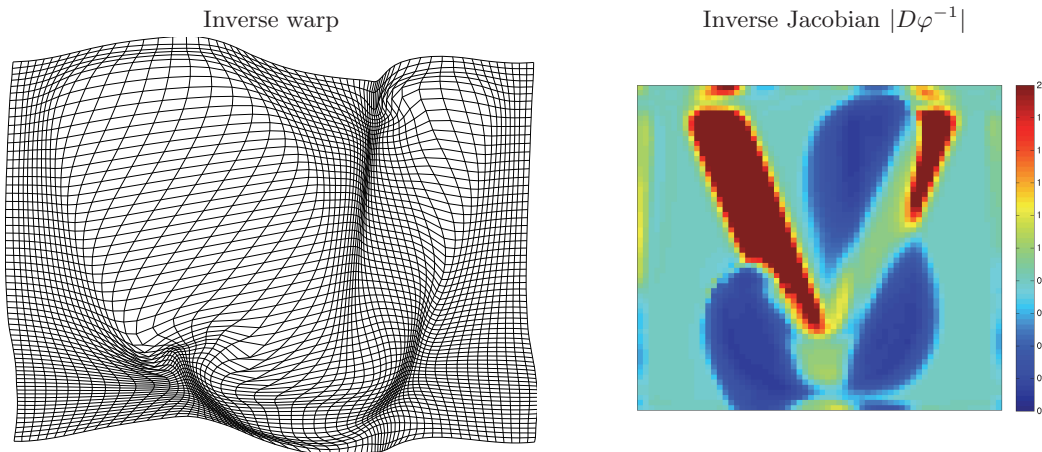We now draw 10^5 uniform samples and transform them with the computed φ:

A benefit of transport-based methods over traditional Markov chain Monte Carlo methods is cheap computation of additional samples; it amount to drawing uniform samples and then evaluating the transformation. A downside is poor scaling with increasing dimensions.

Detailed comparisons of the OIT approach developed here, to the OMT approach developed by Moselhy and Marzouk [29] and Reich [34], is beyond the scope of this paper. Since OIT is intrinsically simpler than OMT (Poisson instead of Monge–Ampere equation), we expect OIT- to outperform OMT-based approaches.

4.2. Registration of letters: J to V. This example illustrates and explains why OIT, i.e., Problem 4, is *not* suitable for image registration. Recall that the algorithm developed for Problem 4 works for any background metric. Thus, given a source density $\mu_0 \in \mathrm{Dens}(M)$, we can *construct* a background metric $\mathbf{g}$ such that $\mathrm{vol}_{\mathbf{g}} = \mu_0$ (various ways of doing this are explained in Appendix A). Suppose now we want to match the letters J and V, represented as gray-scale functions I_0 and I_1 on $M = \mathbb{T}^2$. We might have to add some background density for black pixels, since I_0 and I_1 must be strictly positive in order for Algorithm 1 to be well-defined. Then we construct the *conformal background metric* $\mathbf{g}$ such that $\mathrm{vol}_{\mathbf{g}} = \mu_0$. With step size $\varepsilon = 0.05$ (20 steps) and background density 0.2 (lowest grey-scale value, white corresponds to 1.0), we get the following sequence of warps:

Source   Target

This is simply blending between the images, as foreseen by the formula (6) for Fisher–Rao geodesics. The corresponding mesh deformation and Jacobian of the inverse at the final point look as follows:

Inverse warp   Inverse Jacobian $|D\varphi^{-1}|$

This is *not* a satisfactory registration: instead of transporting the white pixels of the J to the white pixels of the V, the resulting diffeomorphism produces white pixels by compressing the background pixels. This example shows that, although Problem 4 allows matching of any pair

of densities by using a compatible metric, only those applications where μ_0 is the standard volume form are likely to be of interest. A remedy for more general, noncompatible matching applications is developed next.

5. Matching with noncompatible background metric. In section 3 we derived an algorithm for solving Problems 1 and 2 for the case when the background metric g is compatible with μ_0. In this section we want to derive an algorithm for the situation of a noncompatible background metric. When the background metric g is noncompatible, the solution to Problem 2 with respect to the information metric (4) on $\mathrm{Diff}(M)$ and Fisher–Rao metric (3) on $\mathrm{Dens}(M)$ is still obtained by a geodesic curve $\varphi(t)$. However, $\varphi(t)$ is not horizontal and therefore does not project to a geodesic on the space of densities. As a consequence, the main ingredient of our efficient lifting algorithm—the explicit formula for geodesics on $\mathrm{Dens}(M)$— cannot be used. From a geometric standpoint, the problem is that the information metric G^I does not descend to $\mathrm{Diff}_{\mu_0}(M)\backslash\mathrm{Diff}(M)$ unless $\mu_0 = \mathrm{vol}$.

To numerically solve the density matching problem using the LDDMM techniques developed in [7] is plausible, but computationally expensive. In the following, we shall instead study a slightly modified matching problem, for which efficient algorithms can still be obtained.

5.1. Inexact matching with the divergence metric. The modification resides in exchanging the information metric G^I for the degenerate divergence metric G^{div}. The degeneracy of the divergence metric is characterized by

$$\mathrm{div}(U \circ \varphi^{-1}) = 0 \iff G_\varphi^{\mathrm{div}}(U,U) = 0,$$

so the kernel is given by the tangent directions of the fibers of the principal bundle (10), explained in subsection 2.4. As mentioned in the introduction, the divergence metric descends to the Fisher–Rao metric. If d_{div} denotes the distance function of G^{div}, we have

$$(18) \qquad d_{\mathrm{div}}(\mathrm{id}, \varphi) = \bar{d}_F(\mathrm{vol}, \varphi^*\mathrm{vol}) = \bar{d}_F(\varphi_*\mathrm{vol}, \mathrm{vol}).$$

In consequence, the inexact density matching problem with G^{div} and $\bar{G}^F$ is the following.

Problem 6 (inexact matching with divergence metric). *Given* $\mu_0, \mu_1 \in \overline{\mathrm{Dens}}(M)$, *find* $\varphi \in \mathrm{Diff}(M)$ *minimizing*

$$(19) \qquad E(\varphi) = E(\varphi; \mu_0, \mu_1) = \sigma\bar{d}_F^2(\varphi_*\mathrm{vol}, \mathrm{vol}) + \bar{d}_F^2(\varphi_*\mu_0, \mu_1),$$

where $\sigma > 0$ *is a regularization parameter, penalizing change of volume.*

Notice that we allow the source and target densities to belong to the completion $\overline{\mathrm{Dens}}(M)$. This relaxation is possible because of (18) and the fact that the action of $\mathrm{Diff}(M)$ extends naturally from $\mathrm{Dens}(M)$ to $\overline{\mathrm{Dens}}(M)$. For applications of Problem 6 the relaxation is important, as it allows us to treat images as densities. This only works for inexact matching, since Moser's lemma requires strictly positive densities.

Due to the degeneracy of the divergence metric, a solution to Problem 6 is not unique. Indeed, with

$$\mathrm{Diff}_{\mathrm{vol},\mu_0}(M) = \mathrm{Diff}_{\mathrm{vol}}(M) \cap \mathrm{Diff}_{\mu_0}(M),$$
$$\mathrm{Diff}_{\mathrm{vol},\mu_1}(M) = \mathrm{Diff}_{\mathrm{vol}}(M) \cap \mathrm{Diff}_{\mu_1}(M),$$

we have the following result.

Lemma 5.1. *Let $\varphi \in \mathrm{Diff}(M)$, $\eta_0 \in \mathrm{Diff}_{\mathrm{vol},\mu_0}(M)$, and $\eta_1 \in \mathrm{Diff}_{\mathrm{vol},\mu_1}(M)$. Then*

$$E(\eta_1 \circ \varphi) = E(\varphi) = E(\varphi \circ \eta_0).$$

Proof. Since $\eta_1 \in \mathrm{Diff}_{\mathrm{vol},\mu_1}(M)$ we also have that $\eta_1^{-1} \in \mathrm{Diff}_{\mathrm{vol},\mu_1}(M)$, since the intersection of two groups is again a group. Using the invariance of the Fisher–Rao metric we get

$$
\begin{aligned}
E(\eta_1 \circ \varphi) &= \sigma \bar{d}_F^2((\eta_1)_* \varphi_* \mathrm{vol}, \mathrm{vol}) + \bar{d}_F^2((\eta_1)_* \varphi_* \mu_0, \mu_1) \\
&= \sigma \bar{d}_F^2(\varphi_* \mathrm{vol}, \eta_1^* \mathrm{vol}) + \bar{d}_F^2(\varphi_* \mu_0, \eta_1^* \mu_1) = E(\varphi).
\end{aligned}
$$

Again, using the invariance of the Fisher–Rao metric, we get

$$
\begin{aligned}
E(\varphi \circ \eta_0) &= \sigma \bar{d}_F^2(\varphi_*(\eta_0)_* \mathrm{vol}, \mathrm{vol}) + \bar{d}_F^2(\varphi_*(\eta_0)_* \mu_0, \mu_1) \\
&= \sigma \bar{d}_F^2(\varphi_* \mathrm{vol}, \mathrm{vol}) + \bar{d}_F^2(\varphi_* \mu_0, \mu_1) = E(\varphi).
\end{aligned}
$$

This proves the lemma. ∎

Thus, the functional E has two different descending properties:

1. It descends to a functional on the right cosets $\mathrm{Diff}_{\mathrm{vol},\mu_1}(M) \backslash \mathrm{Diff}(M)$, right invariant with respect to $\mathrm{Diff}_{\mathrm{vol},\mu_0}(M)$. The corresponding right principal bundle is

(20)
$$\mathrm{Diff}_{\mathrm{vol},\mu_1}(M) \hookrightarrow \mathrm{Diff}(M)$$
$$\downarrow$$
$$\mathrm{Diff}_{\mathrm{vol},\mu_1}(M) \backslash \mathrm{Diff}(M).$$

2. It descends to a functional on the left cosets $\mathrm{Diff}(M)/\mathrm{Diff}_{\mathrm{vol},\mu_0}(M)$, left invariant with respect to $\mathrm{Diff}_{\mathrm{vol},\mu_1}(M)$. The corresponding left principal bundle is

(21)
$$\mathrm{Diff}_{\mathrm{vol},\mu_0}(M) \hookrightarrow \mathrm{Diff}(M)$$
$$\downarrow$$
$$\mathrm{Diff}(M)/\mathrm{Diff}_{\mathrm{vol},\mu_0}(M).$$

We need a strategy to tackle the degeneracy problem explained in Lemma 5.1. Our approach is simple: we impose the additional constraint on φ that it should be connected to the identity by a curve that is G^I-orthogonal to the fibers of both principle bundles (20) and (21). Then Problem 6 can be solved efficiently by a gradient flow, as we now explain.

5.2. Gradient flow for inexact matching. Let $\nabla^I E$ denote the gradient of E with respect to the information metric G^I. Our approach for Problem 6, i.e., for minimizing the functional E in (19), is to discretize the gradient flow

(22)
$$\dot{\varphi} = -\nabla^I E(\varphi).$$

Since the functional E is constant along the fibers of the principal bundles (20) and (21), the curve traced out by the gradient flow is G^I-orthogonal to both fibers, as desired.

Through formulas (7) and (19) we obtain an explicit formula for E. We can then derive the gradient $\nabla^I E$. It is convenient to carry out the calculations using the sphere representation for the densities via the W-map (5).

Proposition 5.2. *The G^I–gradient of the divergence-metric matching functional E in (19) is given by*

$$
(23) \quad \nabla^I E(\varphi) = A^{-1} \left(\sigma c \left(\int_M \sqrt{|D\varphi^{-1}|}\, \mathrm{vol} \right) \mathrm{grad}\, \sqrt{|D\varphi^{-1}|} \right.
$$
$$
\left. + c \left(\int_M W_\varphi W_1 \mathrm{vol} \right) \left(W_1 \, \mathrm{grad}\, W_\varphi - W_\varphi \, \mathrm{grad}\, W_1 \right) \right) \circ \varphi ,
$$

where A is the inertia operator (9), $W_1 := W(\mu_1)$, $W_\varphi := W(\varphi_* \mu_0)$, and

$$
c \colon [0, \mathrm{vol}(M)] \to \mathbb{R}, \quad c(x) = \frac{\arccos\left(\frac{x}{\mathrm{vol}(M)} \right)}{\sqrt{1 - \frac{x^2}{\mathrm{vol}(M)^2}}} .
$$

Proof. Take a curve $\varphi(\epsilon)$ such that $\varphi(0) = \varphi$, i.e., a variation of φ. Then

(24)
$$
\left. \frac{\mathrm{d}}{\mathrm{d}\epsilon} \right|_{\epsilon=0} E(\varphi(\epsilon)) = G^I_\varphi (\nabla_G E, \dot\varphi(0))
$$
$$
= \sigma \left\langle \frac{\delta \bar{d}_F^2}{\delta\mu}(\varphi_* \mathrm{vol}, \mathrm{vol}), \left. \frac{\mathrm{d}}{\mathrm{d}\epsilon} \right|_{\epsilon=0} \varphi(\epsilon)_* \mathrm{vol} \right\rangle + \left\langle \frac{\delta \bar{d}_F^2}{\delta\mu}(\varphi_* \mu_0, \mu_1), \left. \frac{\mathrm{d}}{\mathrm{d}\epsilon} \right|_{\epsilon=0} \varphi(\epsilon)_* \mu_0 \right\rangle .
$$

We now write the variation of the form $\dot\varphi(\epsilon) = v \circ \varphi(\epsilon)$ for some vector field $v \in \mathfrak{X}(M)$. For any $\mu \in \mathrm{Dens}(M)$ we then have

$$
\left. \frac{\mathrm{d}}{\mathrm{d}\epsilon} \right|_{\epsilon=0} \varphi(\epsilon)_* \mu = -\mathcal{L}_v \varphi_* \mu .
$$

Let $r := \sqrt{\mu(M)} = \sqrt{\nu(M)}$. Then, from (28)

$$
\bar{d}_F^2(\mu, \nu) = r^2 \arccos \left(\frac{1}{r^2} \int_M W(\mu) W(\nu) \mathrm{vol} \right)^2 ,
$$

so the variational derivative is given by

$$
\left\langle \frac{\delta \bar{d}_F^2}{\delta\mu}(\mu, \nu), \eta \right\rangle = \underbrace{\frac{\arccos\left(\frac{1}{r^2} \int_M W(\mu) W(\nu) \mathrm{vol} \right)}{\sqrt{1 - \left(\frac{1}{r^2} \int_M W(\mu) W(\nu) \mathrm{vol} \right)^2}}}_{c(W(\mu) W(\nu))} \left\langle -2 W(\nu) \frac{\delta W(\mu)}{\delta\mu}, \eta \right\rangle .
$$

Since $\langle \frac{\delta W(\mu)}{\delta\mu}, \eta \rangle = \frac{1}{2W(\mu)} \frac{\eta}{\mathrm{vol}}$ we get

$$
\left\langle \frac{\delta \bar{d}_F^2}{\delta\mu}(\mu, \nu), \eta \right\rangle = -c(W(\mu) W(\nu)) \int_M \underbrace{\frac{W(\nu)}{W(\mu)}}_{F(\mu, \nu)} \eta .
$$

Notice that (i) $F(\mu, \mu) = 0$ since the variation η preserves the total volume, (ii) $\varphi_* F(\mu, \nu) = F(\varphi_*\mu, \varphi_*\nu)$ reflecting the invariance of the Fisher–Rao distance, (iii) $\lim_{r\to\infty} c(W(\mu)W(\nu)) = \frac{\pi}{2}$ reflecting the simplified formula when the volume is infinite, and (iv) $c(W(\varphi_*\mathrm{vol})W(\mathrm{vol})) = c(\sqrt{|D\varphi^{-1}|})$. The first term in (24) now becomes

(25)
$$\sigma c(\sqrt{|D\varphi^{-1}|}) \left\langle \frac{\delta \bar{d}_F^2}{\delta \mu}(\varphi_*\mathrm{vol}, \mathrm{vol}), \frac{\mathrm{d}}{\mathrm{d}\epsilon}\bigg|_{\epsilon=0} \varphi(\epsilon)_*\mathrm{vol} \right\rangle$$

$$= \sigma c(\sqrt{|D\varphi^{-1}|}) \int_M F(\varphi_*\mathrm{vol}, \mathrm{vol})\mathcal{L}_v\varphi_*\mathrm{vol}$$

$$= \sigma c(\sqrt{|D\varphi^{-1}|}) \int_M F(\varphi_*\mathrm{vol}, \mathrm{vol})\mathrm{di}_v\varphi_*\mathrm{vol}$$

$$= -\sigma c(\sqrt{|D\varphi^{-1}|}) \int_M \mathrm{d}F(\varphi_*\mathrm{vol}, \mathrm{vol}) \wedge W(\varphi_*\mathrm{vol})^2 \mathrm{i}_v\mathrm{vol}$$

$$= -\sigma c(\sqrt{|D\varphi^{-1}|}) \int_M \mathrm{d}\left(\frac{1}{W(\varphi_*\mathrm{vol})}\right) \wedge W(\varphi_*\mathrm{vol})^2 \mathrm{i}_v\mathrm{vol}$$

$$= \sigma c(\sqrt{|D\varphi^{-1}|}) \int_M \mathrm{d}W(\varphi_*\mathrm{vol}) \wedge \mathrm{i}_v\mathrm{vol}$$

$$= \sigma c(\sqrt{|D\varphi^{-1}|}) \int_M \mathbf{g}(AA^{-1}\,\mathrm{grad}(W(\varphi_*\mathrm{vol})), v)\mathrm{vol}$$

$$= G_{\mathrm{id}}^I\left(\sigma c(\sqrt{|D\varphi^{-1}|})A^{-1}\,\mathrm{grad}(W(\varphi_*\mathrm{vol})), \dot{\varphi}\circ\varphi^{-1}\right)$$

$$= G_\varphi^I\left(\sigma c(\sqrt{|D\varphi^{-1}|})A^{-1}\,\mathrm{grad}\sqrt{|D\varphi^{-1}|}\circ\varphi, \dot{\varphi}\right).$$

Using the notation $W_\varphi := W(\varphi_*\mu_0)$ and $W_1 := W(\mu_1)$, the second term in (24) becomes

(26)
$$\left\langle \frac{\delta \bar{d}_F^2}{\delta \mu}(\varphi_*\mu_0, \mu_1), \frac{\mathrm{d}}{\mathrm{d}\epsilon}\bigg|_{\epsilon=0} \varphi(\epsilon)_*\mu_0 \right\rangle$$

$$= c(W_\varphi W_1) \int_M F(\varphi_*\mu_0, \mu_1)\mathcal{L}_v\varphi_*\mu_0$$

$$= c(W_\varphi W_1) \int_M F(\varphi_*\mu_0, \mu_1)\mathrm{di}_v\varphi_*\mu_0$$

$$= -c(W_\varphi W_1) \int_M \mathrm{d}F(\varphi_*\mu_0, \mu_1) \wedge W_\varphi^2 \mathrm{i}_v\mathrm{vol}$$

$$= -c(W_\varphi W_1) \int_M \mathrm{d}\left(\frac{W_1}{W_\varphi}\right) \wedge W_\varphi^2 \mathrm{i}_v\mathrm{vol}$$

$$= c(W_\varphi W_1) \int_M (W_1\mathrm{d}W_\varphi - W_\varphi\mathrm{d}W_1) \wedge \mathrm{i}_v\mathrm{vol}$$

$$= c(W_\varphi W_1) \int_M \mathbf{g}\left(AA^{-1}(W_1\,\mathrm{grad}\,W_\varphi - W_\varphi\,\mathrm{grad}\,W_1), v\right)\mathrm{vol}$$

$$= G_\varphi^I\left(c(W_\varphi W_1)A^{-1}(W_1\,\mathrm{grad}\,W_\varphi - W_\varphi\,\mathrm{grad}\,W_1)\circ\varphi, \dot{\varphi}\right).$$

Put together, (25) and (26) prove formula (23). ∎

Based on Proposition 5.2, we can now discretize the gradient flow (22). Indeed, a numerical method is given by the following algorithm.

ALGORITHM 2. *Assume we have a numerical way to represent functions, vector fields, and diffeomorphisms on M, and numerical methods for* (i) *composing diffeomorphisms with functions,* (ii) *composing diffeomorphisms with diffeomorphisms,* (iii) *computing the divergence of a vector field,* (iv) *computing the gradient of a function, and* (v) *computing the inverse of the inertia operator A in* (9). *A computational algorithm for the gradient flow* (22) *is then given as follows.*

1. *Choose a step size $\varepsilon > 0$ and initialize the diffeomorphisms $\varphi_0 = \mathrm{id}$, $\varphi_0^{-1} = \mathrm{id}$, and a function $J_0 = 1$. Precompute $W(\mu_1)$ and $\mathrm{grad}(W(\mu_1))$. Set $k \leftarrow 0$.*
2. *Compute action on the source*

$$f_k = W(\mu_0) \circ \varphi_k^{-1} \sqrt{J_k}.$$

3. *Compute coefficients $a_k = c\left(\int_M \sqrt{J_k}\,\mathrm{vol}\right)$ and $b_k = c\left(\int_M f_k W(\mu_1)\mathrm{vol}\right)$, then compute momentum*

$$m_k = \sigma a_k \,\mathrm{grad}(\sqrt{J_k}) + b_k W(\mu_1) \,\mathrm{grad}\, f_k - b_k f_k \,\mathrm{grad}(W(\mu_1)).$$

(If $\mathrm{vol}(M) = \infty$, set $a_k = b_k = 1$.)
4. *Compute the vector field infinitesimally generating the negative gradient*

$$v_k = -A^{-1} m_k.$$

5. *Construct an approximation ψ_k^{-1} to $\exp(-\varepsilon v_k)$, for example, $\psi_k^{-1} = \mathrm{id} - \varepsilon v_k$.*
6. *Update the inverse diffeomorphism: $\varphi_{k+1}^{-1} = \varphi_k^{-1} \circ \psi_k^{-1}$.*
7. *Update the inverse Jacobian using Lie–Trotter splitting:[1]*

$$J_{k+1} = (J_k \circ \psi_k^{-1}) \mathrm{e}^{-\varepsilon \,\mathrm{div}(v_k)}.$$

8. *Construct an approximation ψ_k to $\exp(\varepsilon v_k)$, for example, $\psi_k = \mathrm{id} + \varepsilon v_k$ (output).*
9. *Update the forward diffeomorphism: $\varphi_{k+1} = \psi_k \circ \varphi_k$ (output).*
10. *Set $k \leftarrow k + 1$ and continue from step 2.*

5.3. Geometry of the gradient flow. In this section we describe the geometry associated with the divergence-metric functional E in (19) and the corresponding gradient flow (22).

The diffeomorphism group acts diagonally on $\mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$ by $\varphi \cdot (\mu, \nu) = (\varphi_* \mu, \varphi_* \nu)$. The isotropy group of $(\mu, \nu) \in \mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$, i.e., the subgroup of $\mathrm{Diff}(M)$ that leaves (μ, ν) invariant, is given by

$$\mathrm{Diff}_{\mu,\nu}(M) := \mathrm{Diff}_\mu(M) \cap \mathrm{Diff}_\nu(M).$$

The action of $\mathrm{Diff}(M)$ on $\mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$ is *not* transitive, so there is more than one group orbit. The group orbit through (vol, μ_0), given by

$$\mathrm{Orb}(\mathrm{vol}, \mu_0) := \mathrm{Diff}(M) \cdot (\mathrm{vol}, \mu_0) \subset \mathrm{Dens}(M) \times \mathrm{Dens}(M),$$

---

[1]Here, one could use Strang splitting to obtain higher order.

is a way to represent the quotient set $\mathrm{Diff}(M)/\mathrm{Diff}_{\mathrm{vol},\mu_0}(M)$. This set is potentially complicated (an orbifold), but let us assume we stay away from singular points so we can work with $\mathrm{Orb}(\mathrm{vol},\mu_0)$ as a submanifold of $\mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$. The principal bundle (21) can then be represented as

$$(27) \qquad \begin{array}{c} \mathrm{Diff}_{\mathrm{vol},\mu_0}(M) \lhook\joinrel\longrightarrow \mathrm{Diff}(M) \\ \downarrow \scriptstyle{\pi_{\mathrm{vol},\mu_0}} \\ \mathrm{Orb}(\mathrm{vol},\mu_0)\,, \end{array}$$

where

$$\pi_{\mathrm{vol},\mu_0}(\varphi) = \varphi \cdot (\mathrm{vol},\mu_0) = (\varphi_*\mathrm{vol}, \varphi_*\mu_0).$$

The manifold $\mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$ comes with a metric, namely,

$$\bar{G}^{FF}_{(\mu,\nu)}\big((\alpha_\mu, \alpha_\nu), (\beta_\mu, \beta_\nu)\big) = \sigma \bar{G}^F_\mu(\alpha_\mu, \beta_\mu) + \bar{G}^F_\nu(\alpha_\nu, \beta_\nu).$$

(Notice that $\bar{G}^F$ is naturally extended to a metric on $\overline{\mathrm{Dens}}(M)$ by the W-map (5).) The corresponding distance is given by

$$(28) \qquad \bar{d}^2_{FF}((\mu_1,\nu_1),(\mu_2,\nu_2)) = \sigma \bar{d}^2_F(\mu_1,\mu_2) + \bar{d}^2_F(\nu_1,\nu_2).$$

To connect back to the divergence-metric matching in Problem 6, the key observation is that

$$E(\varphi) = \bar{d}^2_{FF}((\mathrm{vol},\mu_1),(\mathrm{vol},\mu_0)\cdot\varphi) = \bar{d}^2_{FF}((\mathrm{vol},\mu_1), \pi_{\mathrm{vol},\mu_0}(\varphi)),$$

so $E(\varphi)$ is simply the function $(\mu,\nu) \mapsto \bar{d}^2_{FF}((\mathrm{vol},\mu_1),(\mu,\nu))$ on $\mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$ lifted to $\mathrm{Diff}(M)$.

Let us now discuss how the metric G^I fits into this geometry.

Lemma 5.3. *The information metric G^I on $\mathrm{Diff}(M)$ is descending with respect to the principal bundle* (27), *i.e., it descends to a metric $\bar{G}^{\mathrm{Orb}}$ on $\mathrm{Orb}(\mathrm{vol},\mu_0)$.*

Proof. Translation by $\psi \in \mathrm{Diff}_{\mathrm{vol},\mu_0}(M)$ of a vector $U \in T_\varphi\mathrm{Diff}(M)$ along the fiber of (27) is given by $U \mapsto U \circ \psi$. Therefore, the metric G^I is descending if and only if

$$(29) \qquad G^I_\varphi(U,V) = G^I_{\varphi\circ\psi}(U\circ\psi, V\circ\psi) \quad \forall U,V \in \mathcal{H}^{\mathrm{vol},\mu_0}_\varphi,$$

where $\mathcal{H}^{\mathrm{vol},\mu_0}_\varphi$ is the distribution that is G^I-orthogonal to the tangent spaces of the fibers of (27). Since G^I is right invariant, condition (29) is automatically true. ∎

Although G^I descends to a metric $\bar{G}^{\mathrm{Orb}}$, the associated distance function $\bar{d}_{\mathrm{Orb}}$ is *not* explicitly computable, in particular it is not given by $\bar{d}_{FF}$. If it was, the result in Lemma 5.3 would allow us to use the same technique as in section 3 to solve the noncompatible, nondegenerate matching problem using G^I and $\bar{G}^F$, namely, to lift the geodesics on $\mathrm{Orb}(\mathrm{vol},\mu_0)$. Our remedy is to exchange $\bar{d}_{\mathrm{Orb}}$ for the distance function $\bar{d}_{FF}$ on the ambient manifold $\mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$. The geometry of our setup is illustrated in Figure 1.
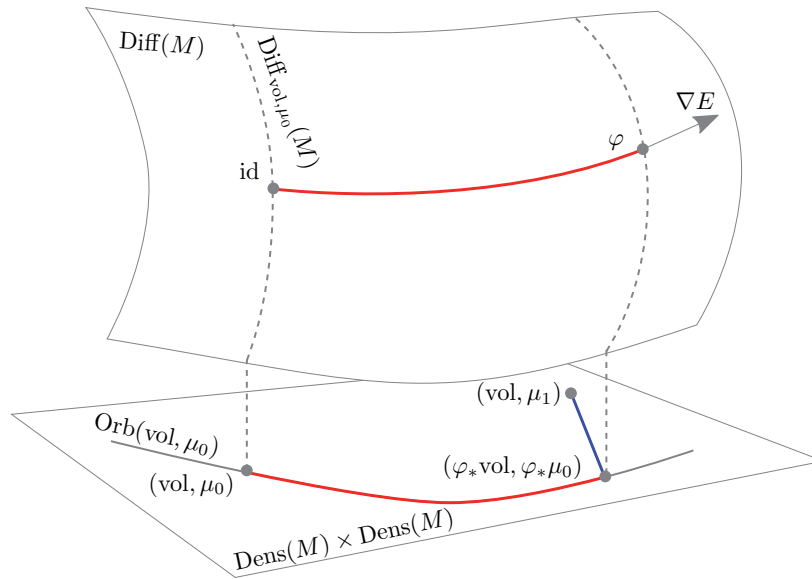
Figure 1. *Illustration of the geometry associated with inexact density matching using the divergence metric. The gradient flow on $\mathrm{Diff}(M)$ descends to a gradient flow on the orbit $\mathrm{Orb}(\mathrm{vol}, \mu_0)$. While constrained to $\mathrm{Orb}(\mathrm{vol}, \mu_0) \subset \mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$, this flow strives to minimize the ambient Fisher–Rao distance $\bar{d}_{FF}$ to (vol, μ_1).*

5.4. Two-component gradient flow. Since both the information metric G^I and the functional E descend with respect to (27), the gradient flow (22) induces a gradient flow on $\mathrm{Orb}(\mathrm{vol}, \mu_0) \subset \mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$. This allows us to represent the gradient flow on $\mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$.

Proposition 5.4. *The gradient flow (22) descends to a two-component gradient flow equation on $\mathrm{Dens}(M) \times \overline{\mathrm{Dens}}(M)$, constrained to stay on $\mathrm{Orb}(\mathrm{vol}, \mu_0)$. Expressed in the variables $J = |D\varphi^{-1}|$ and $P = (\varphi_*\mu_0)/\mathrm{vol}$ it is given by*

$$(30) \qquad \begin{aligned} \dot{J} &= -v \cdot \mathrm{grad}\, J - J \,\mathrm{div}(v), \\ \dot{P} &= -v \cdot \mathrm{grad}\, P - P \,\mathrm{div}(v) \end{aligned}$$

with

$$v = A^{-1}\left(\sigma c \left(\int_M \sqrt{J}\,\mathrm{vol} \right) \mathrm{grad}\,\sqrt{J} \right.$$
$$\left. + c \left(\int_M W(\mu_1)\sqrt{P}\,\mathrm{vol} \right) \left(W(\mu_1)\,\mathrm{grad}\,\sqrt{P} - \sqrt{P}\,\mathrm{grad}\,W(\mu_1) \right) \right).$$

Proof. Let $\varphi(t)$ be the solution of the gradient flow (22). Then $\dot{\varphi}(t) \circ \varphi(t)^{-1} = \nabla^I E(\varphi(t)) \circ \varphi(t)^{-1} =: v(t)$ depends on $\varphi(t)$ only through $\varphi(t)_*\mathrm{vol}$ and $\varphi(t)_*\mu_0$, i.e., through J and P. We also have

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t)_*\mathrm{vol} = -\mathcal{L}_{v(t)}\varphi(t)_*\mathrm{vol} = -\left(\mathcal{L}_{v(t)} J + \mathrm{div}(v(t))J \right)\mathrm{vol}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t)_*\mu_0 = -\mathcal{L}_{v(t)}\varphi(t)_*\mu_0 = -\left(\mathcal{L}_{v(t)}P + \mathrm{div}(v(t))P\right)\mathrm{vol}.$$

This proves the result. ∎

This proposition tells us that an alternative way to compute the gradient flow (22) is to first solve (30) and then the lifting equations for the principle bundle (27), thereby obtaining a horizontal curve in $\mathrm{Diff}(M)$. We have not investigated this approach in detail.

5.5. Relation to matching with compatible background metric. In this section we want to show the relation between the compatible background approach in section 3 and the gradient flow approach developed here. Recall that the solutions to Problems 4 and 6 are obtained by the inverse of the endpoint of a horizontal geodesic, obtained by lifting a Fisher–Rao geodesic. As a consequence, we have the following two results.

Lemma 5.5. *Let $\mu_0 = \mathrm{vol}, \mu_1 \in \mathrm{Dens}(M)$, and let $[0,1] \mapsto \gamma(t) \in \mathrm{Diff}(M)$ be the horizontal G^I-geodesic such that $\gamma(1)^{-1}$ solves the OIT problem (4). Then $\nabla E(\gamma(t); \mu_1, \mathrm{vol})$ is parallel with $\dot{\gamma}(t)$.*

Proof. We have $E(\varphi; \mu_1, \mathrm{vol}) = \sigma \bar{d}_F^2(\varphi^*\mathrm{vol}, \mathrm{vol}) + \bar{d}_F^2(\varphi^*\mathrm{vol}, \mu_1)$. Thus, $E(\cdot; \mu_1, \mathrm{vol})$ descends with respect to Moser's principal bundle (10). Since the information metric G^I descends to the Fisher–Rao metric, the gradient flow descends to a gradient flow on $\mathrm{Dens}(M)$, given by

$$(31) \qquad\qquad \dot{\mu} = -\nabla^F e(\mu), \quad e(\mu) = \sigma\bar{d}_F^2(\mathrm{vol}, \mu) + \bar{d}_F^2(\mu, \mu_1),$$

where ∇^F is the gradient with respect to $\bar{G}^F$. If now $\mu = \pi(\gamma(t))$, then $\nabla^F e(\mu)$ is parallel with $\frac{\mathrm{d}}{\mathrm{d}t}\pi(\gamma(t))$, since $\pi(\gamma(t))$ is the unique minimizing geodesic between vol and μ_1. The result now follows since two horizontally lifted paths are parallel if they are parallel on $\mathrm{Dens}(M)$. ∎

Proposition 5.6. *If $\mu_0 = \mathrm{vol}$ and $\mu_1 \in \mathrm{Dens}(M)$, then the limit of the gradient flow $\dot{\varphi} = -\nabla_A E(\varphi; \mu_1, \mathrm{vol})$, $\varphi(0) = \mathrm{id}$, exists and coincides with the inverse of the solution to the nonexact OIT problem (5).*

Proof. The proof follows since the gradient flow $\dot{\varphi} = -\nabla_A E(\varphi; \mu_1, \mathrm{vol})$ descends to the gradient flow (31) and $e(\mu)$ is a strictly convex functional and $\mathrm{Dens}(M)$ is a convex space with respect to the Fisher–Rao geometry. ∎

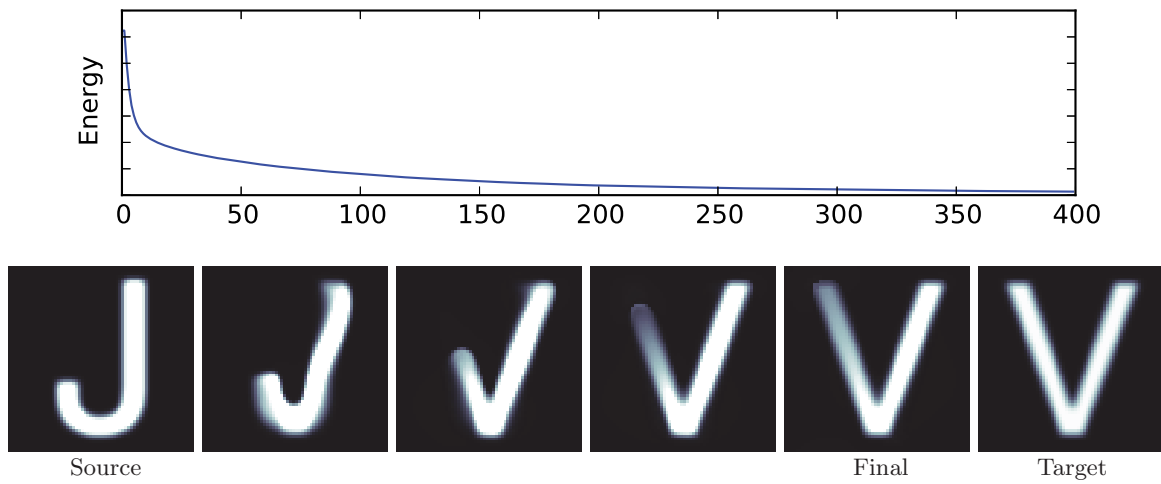*Remark* 5.7. In contrast to the previous parts of this section we actually require here that μ_1 is strictly positive. This is indeed a necessary condition: for μ_1 on the boundary of $\mathrm{Dens}(M)$ we cannot guarantee the existence of a minimizer to the nonexact OIT problem. In fact for a target density $\mu_1 \in \overline{\mathrm{Dens}}(M)$ the optimal deformation φ will, in general, not be a diffeomorphisms, instead it will have a vanishing derivative on certain points or even intervals. To guarantee the existence of minimizers in this situation also, one could use a complete metric on the diffeomorphism group as regularization term. Possible choices for this include higher order Sobolev metrics [12, 5, 8, 25] and metrics that are induced by Gaussian kernels [36].

6. Examples of matching with noncompatible metric. In this section we evaluate the gradient-flow-based Algorithm 2 in various examples where $\mu_0, \mu_1 \in \overline{\mathrm{Dens}}(\mathbb{T}^2)$, i.e., there are
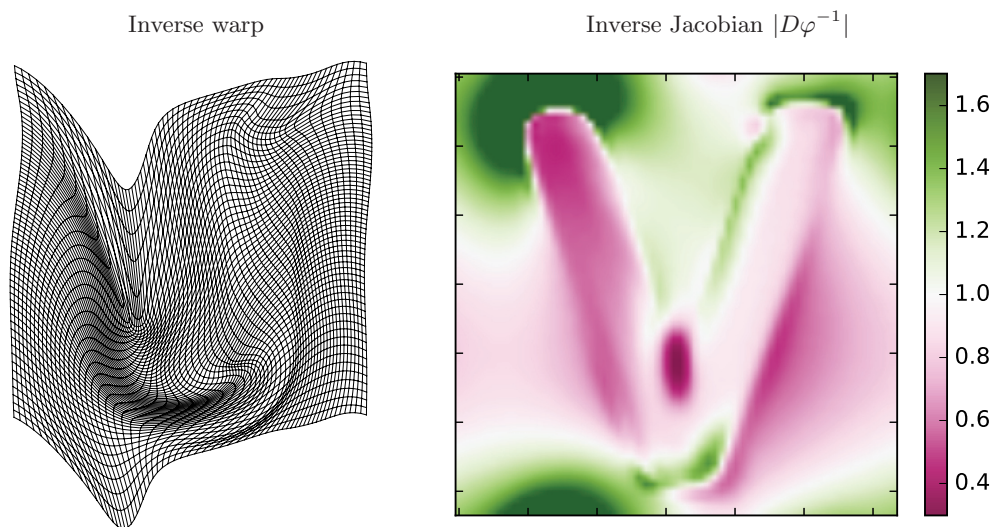
regions with vanishing density (represented by black pixels). For simplicity, in our experiments we consider the flat, periodic torus and use the FFT for inverting the operator A in (9). There has been extensive work on fast, efficient solutions of Poisson's problem on other manifolds; see, for example, the review [11].

6.1. Registration of letters: J to V. This example illustrates the capability of producing severely deformed warps, namely, to warp the letter J into the letter V. We also examine the effect of the balancing parameter σ.

With step size $\varepsilon = 0.2$, balance $\sigma = 0.05$, and 400 iterations, we get the following energy evolution and sequence of warps:
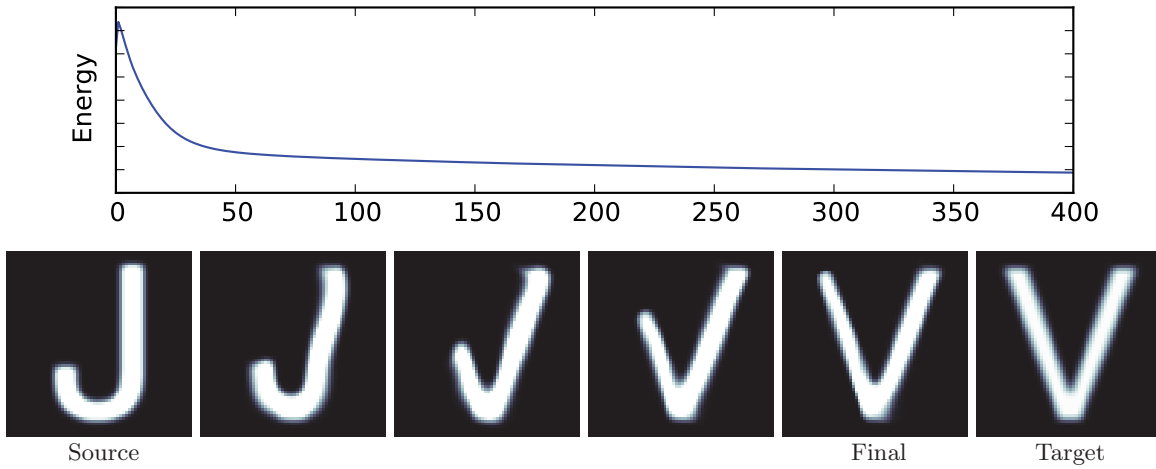

Notice in the warp that the pixel values have changed in regions of large deformations (the upper left part of the final V). This is due to expansion. The corresponding mesh deformation and Jacobian of the inverse at the final point look as follows:
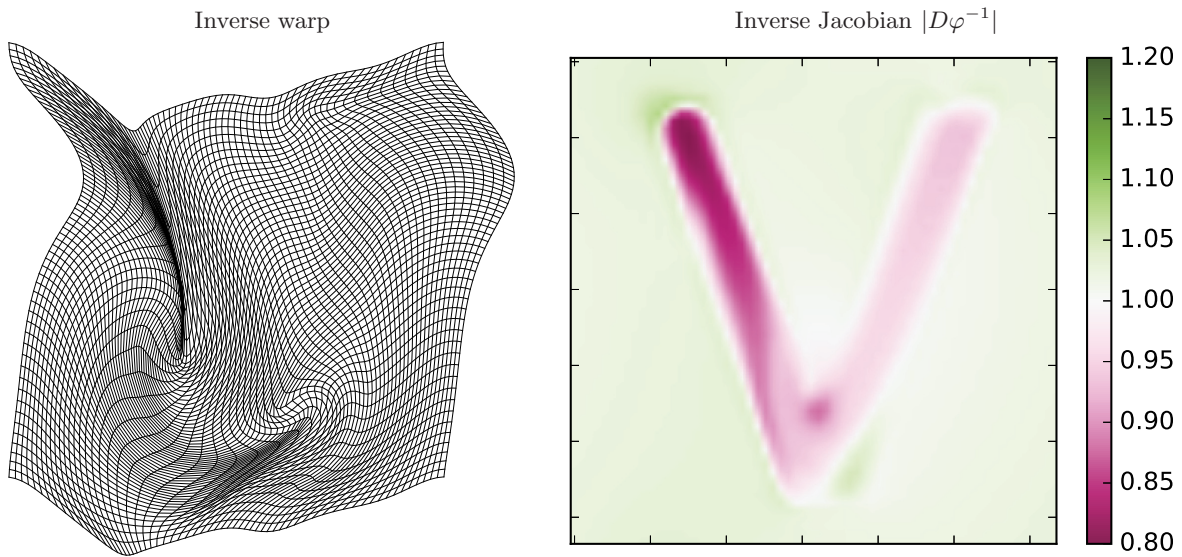

Notice how the lower part of the J is stretched out to the left part of the V.

Let us now do the same experiment but with the larger balancing parameter $\sigma = 5$:
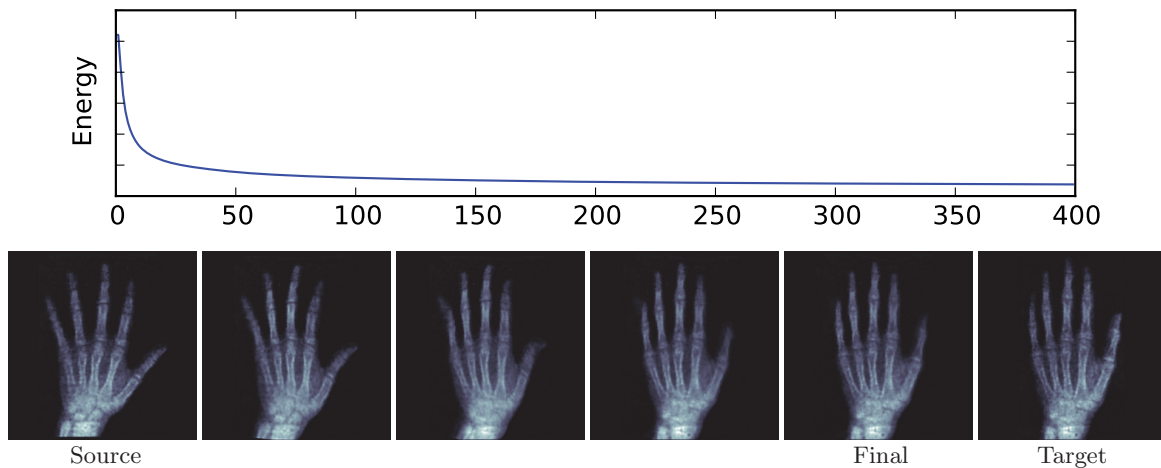

Notice here that the warp rarely changes the pixel values, at the price of less expansion of the left part of the V. This is due to a smaller amount of compression and expansion, as seen below in the corresponding mesh deformation and Jacobian of the inverse:
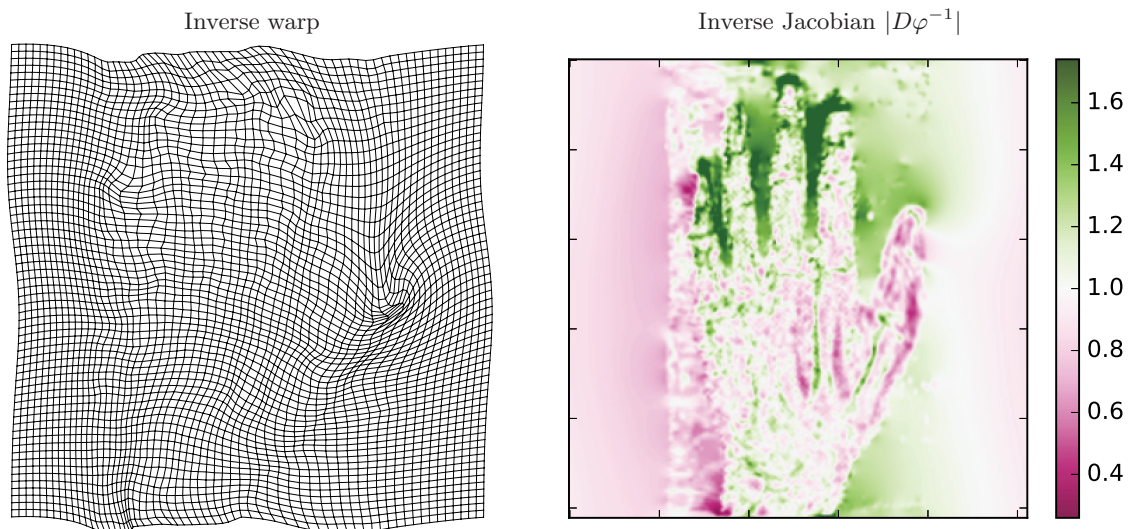

Compared to the smaller σ, the Jacobian determinant is more regular and closer to one and the mesh deformation is almost volume preserving. This example illustrates nicely the influence of the balancing parameter.

6.2. Registration of noisy x-ray images. In this example we illustrate the use of Algorithm 2 for registration of low-resolution, noisy x-ray images of human hands. With step size $\varepsilon = 0.2$, balance $\sigma = 0.1$, and 400 iterations, we get the following energy evolution and sequence of warps:

Source   Final   Target

Except for the tip of the little finger and the thumb, the resulting path of diffeomorphisms yields a good warp between corresponding bone structures in the hands. The mesh deformation and Jacobian of the inverse at the final point look as follows:

Inverse warp   Inverse Jacobian $|D\varphi^{-1}|$

Here one can see how the noise affects the diffeomorphisms: both the mesh warp and the Jacobian are somewhat irregular, except at the left and right borders where the source and target densities vanish.

Appendix A. Constructing compatible background metrics. In section 3 we have described a method to solve the density matching problem, assuming that we are given a compatible background metric. If the source density μ_0 is not equal to the density induced by the background metric, one has to construct such a metric first (as in Example 4.2). There is, of course, a range of background metrics h having a prescribed volume form; here we describe two specific choices.

A.1. Conformal metric. As already discussed, any volume form $\mu \in \mathrm{Dens}(M)$ can be written $\mu = I\mathrm{vol_g}$. Note that I is the Radon–Nikodym derivative of μ with respect to $\mathrm{vol_g}$ (as measures). This observation yields a first choice for the desired metric.

Lemma A.1. *Let $\mu = I\mathrm{vol_g} \in \mathrm{Dens}(M)$. Then the modified metric*

$$\mathsf{h} = I^{\frac{2}{n}}\mathsf{g} \tag{32}$$

is compatible with μ, i.e., $\mathrm{vol_h} = \mu$.

Proof. In coordinates $(x^1, \ldots, x^n)$ we have

$$\mathrm{vol_h} = \sqrt{\det(\mathsf{h}_{ij})}\, dx^1 \wedge \ldots \wedge dx^n. \quad \blacksquare$$

Remark A.2. Note that the metric (32) is conformally equivalent to the background metric g. In mechanics, it is called the *Jacobi metric*.

The advantage of the metric (32) is that it is easy to construct and also that the Laplacian and gradient take simple forms (in terms of the Laplacian and gradient of the original background metric g).

Lemma A.3. *The Laplacian and gradient of the metric $\mathsf{h} = I^{\frac{n}{2}}\mathsf{g}$ are given by*

$$\mathrm{grad_h}(f) = I^{-\frac{n}{2}}\,\mathrm{grad_g}(f), \qquad \Delta_{\mathsf{h}}f = \frac{1}{I}\,\mathrm{div_g}(I^{1-\frac{n}{2}}\,\mathrm{grad_g}(f)).$$

Proof. To calculate the expression for the gradient we use

$$df(X) = \mathsf{h}(\mathrm{grad_h}(f), X) = \mathsf{g}(I^{\frac{n}{2}}\,\mathrm{grad_h}(f), X) = \mathsf{g}(\mathrm{grad_g}(f), X).$$

For the Laplacian we express the divergence in coordinates

$$\mathrm{div}_g(X) = \frac{1}{\sqrt{\det \mathsf{h}_{ij}}}\partial_i(\sqrt{\det \mathsf{h}_{ij}}\, X^i) = \frac{1}{I}\partial_i(I\,X^i) = \frac{1}{I}\,\mathrm{div_g}(IX).$$

Then we have

$$\Delta_{\mathsf{h}}f = \mathrm{div_h}(\mathrm{grad_h}(f)) = \mathrm{div_h}(I^{-\frac{n}{2}}\,\mathrm{grad_g}(f)) = \frac{1}{I}\,\mathrm{div_g}(I \cdot I^{-\frac{n}{2}}\,\mathrm{grad_g}(f))$$

$$= \frac{1}{I}\,\mathrm{div_g}(I^{1-\frac{n}{2}}\,\mathrm{grad_g}(f)) \quad \blacksquare$$

Remark A.4. Note that for $n = 2$ the formula for the Laplacian simplifies to

$$\Delta_{\mathsf{h}} = \frac{1}{I}\Delta_{\mathsf{g}},$$

reflecting the scale invariance of the Laplacian in dimension two.

Corollary A.5. *Using the metric (32) the lifting equation (15) reads*

$$\dot{\varphi}(t) = v(t) \circ \varphi(t), \quad \varphi(0) = \mathrm{id},$$

$$v(t) = I^{-\frac{n}{2}}\,\mathrm{grad_g}(f),$$

$$\frac{1}{I}\,\mathrm{div_g}(I^{1-\frac{n}{2}}\,\mathrm{grad_g}(f)) = \frac{\dot{\mu}(t)}{\mu(t)} \circ \varphi(t)^{-1}.$$

A.2. Constructing a flat compatible metric. For a flat background metric g, the Jacobi metric h given by (32) is not, in general, flat. Indeed, h is only flat if I is constant. Now we will describe a method to choose a flat background metric assuming that the background metric g is flat.

Lemma A.6. *Assume that M carries a flat background metric* g. *Let $\mu \in \mathrm{Dens}(M)$ and let φ be the solution of Problem* 4 *with $\mu_0 = \mathrm{vol}$ and $\mu_1 = \mu$. Then* $h = \varphi^* g$ *is a flat metric and* $\mathrm{vol_h} = \mu$.

Proof. We have $\mathrm{vol}_{\varphi^* h} = \varphi^* \mathrm{vol_h}$ for any metric h and therefore $h = \varphi^* g$ has the desired volume density. The flatness follows since the curvature tensor $\mathcal{R}_g$ of g satisfies $0 = \mathcal{R}_g = \varphi_* \mathcal{R}_{\varphi^* g}$. ∎

A.3. Symmetric matching. In the previous section we have described how to choose a metric that is compatible with a fixed volume form. Thus, the solution of Problem 1 with respect to the Fisher–Rao metric (3) can be obtained as follows:

1. Given $\mu_0, \mu_1 \in \mathrm{Dens}(M)$, construct a background metric h compatible with μ_0 by subsections A.1 or A.2.
2. Solve the lifting equation (15) with respect to h, using the explicit geodesic curve (6) between μ_0 and μ_1.

This problem is not symmetric in μ_0 and μ_1, since the metric h depends on μ_0. However, the condition that h is compatible with μ_0 can be made more general, which allow us to derive a symmetric solution. We use the fact that any geodesic that is horizontal at some $s \in (0, 1)$ remains horizontal for all time. Thus we do not have to choose a metric that is compatible with one of the prescribed densities μ_0 and μ_1, but only a metric that is compatible with some density that lies on the geodesic connecting the two. This suggests the following choice of background metric h, which leads to a symmetric solution of Problem 1:

1. Calculate the geodesic curve $\mu(t)$ connecting μ_0 to μ_1 using (6).
2. Construct a background metric h that is compatible with the midpoint of the geodesic $\mu(\frac{1}{2})$ using subsections A.1 or A.2.
3. Solve the lifting equations for the part of the geodesic that connects the densities $\mu(\frac{1}{2})$ and μ_1 with respect to the compatible metric h. Denote the resulting deformation φ.
4. Solve the lifting equations for the part of the geodesic that connects the densities $\mu(\frac{1}{2})$ and μ_0 with respect to the compatible metric h. Denote the resulting deformation ψ.
5. The solution is then given by $\varphi \circ \psi^{-1}$.

The symmetry of this strategy follows from the construction.

Appendix B. Relation between Fisher–Rao and other distances.

In this section we compare the Fisher–Rao distance $\bar{d}_F(\cdot, \cdot)$ with the Wasserstein distance $d^W(\cdot, \cdot)$, the Hellinger distance $d^H(\cdot, \cdot)$, the total variation distance $d^{TV}(\cdot, \cdot)$, the Kullback–Leibler divergence $d^{KL}(\cdot, \cdot)$, and the χ^2-distance $d^\chi(\cdot, \cdot)$ (see [15] for definitions). Note that the Kullback–Leibler divergence and the χ^2-distance are not metric distances, as they are not symmetric.

The following inequalities for probability distances are given by Gibbs and Su [15]:

1. $\dfrac{d^W(\mu_-, \mu_+)}{\sup_{x,y \in M} \left(d^M(x, y) \right)} \leq d^H(\mu_-, \mu_+) \leq \inf_{x,y \in M} \left(d^M(x, y) \right) d^W(\mu_-, \mu_+),$
2. $d^{TV}(\mu_-, \mu_+) \leq d^H(\mu_-, \mu_+) \leq \sqrt{d^{TV}(\mu_-, \mu_+)},$

3. $d^{\mathrm{H}}(\mu_-, \mu_+) \le \sqrt{d^{\mathrm{KL}}(\mu_-, \mu_+)}$,
4. $d^{\mathrm{H}}(\mu_-, \mu_+) \le \sqrt{d^{\chi}(\mu_-, \mu_+)}$.

Here, M is more general than before: it can be any metric space. The lower bound in item 1, however, is only meaningful for bounded M and the upper bound only for discrete M. Using these inequalities, and a comparison of the Fisher–Rao distance with the Hellinger distance, we obtain the following result.

Proposition B.1. *For any two densities μ_0, $\mu_1 \in \mathrm{Dens}(M)$ we have*

5. $d^{\mathrm{H}}(\mu_-, \mu_+) \le d^{\mathrm{FR}}(\mu_-, \mu_+) \le \frac{\pi}{2} d^{\mathrm{H}}(\mu_-, \mu_+)$,

6. $\dfrac{d^{\mathrm{W}}(\mu_-, \mu_+)}{\sup_{x,y\in M}\left(d^M(x,y)\right)} \le d^{\mathrm{FR}}(\mu_-, \mu_+) \le \dfrac{\pi}{2} \inf_{x,y\in M}\left(d^M(x,y)\right) \, d^{\mathrm{W}}(\mu_-, \mu_+)$,

7. $d^{\mathrm{TV}}(\mu_-, \mu_+) \le d^{\mathrm{FR}}(\mu_-, \mu_+) \le \sqrt{\frac{\pi}{2} d^{\mathrm{TV}}(\mu_-, \mu_+)}$,

8. $d^{\mathrm{FR}}(\mu_-, \mu_+) \le \sqrt{\frac{\pi}{2} d^{\mathrm{KL}}(\mu_-, \mu_+)}$,

9. $d^{\mathrm{FR}}(\mu_-, \mu_+) \le \sqrt{\frac{\pi}{2} d^{\chi}(\mu_-, \mu_+)}$.

Proof. To prove the first inequality we recall that $d^{\mathrm{FR}}(\mu_0, \mu_1)$ equals the spherical Hellinger distance; see subsection 2.2. Thus we only have to compare the spherical distance to the Euclidean distance, yielding a factor $\frac{\pi}{2}$. Now the other inequalities follow immediately using items 1 to 4. ∎

Recall from section 2 that the Fisher–Rao metric arises from a metric on $\mathrm{Diff}(M)$. Likewise, there is a metric on $\mathrm{Dens}(M)$, corresponding to the Wasserstein distance, that is the projection of a (non-right-invariant) metric on the diffeomorphism group [31]. To our knowledge, it is not known whether similar statements exist for the other distance functions above. Notice that the metric on $\mathrm{Dens}(M)$ corresponding to the Wasserstein distance is of Sobolev-type H^{-1}, whereas the Fisher–Rao metric is an L^2-type metric. Likewise, the metric on $\mathrm{Diff}(M)$ inducing the Wasserstein distance is of L^2-type, whereas the information metric is of Sobolev-type H^1. Further discussions about the relation between metrics on $\mathrm{Diff}(M)$ and $\mathrm{Dens}(M)$ are given in [6, Chap. 3].

REFERENCES

[1] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Trans. Math. Monogr. 191, AMS, Providence, RI, 2000.

[2] V. I. Arnold and B. A. Khesin, *Topological Methods in Hydrodynamics*, Appl. Math. Sci. 125, Springer-Verlag, New York, 1998.

[3] M. Bauer, M. Bruveris, and P. Michor, *Homogeneous Sobolev metric of order one on diffeomorphism groups on real line*, J. Nonlinear Sci., 24 (2014), pp. 769–808.

[4] M. Bauer, M. Bruveris, and P. Michor, *Uniqueness of the Fisher–Rao Metric on the Space of Smooth Densities*, preprint, arXiv:1411.5577, 2014.

[5] M. Bauer, J. Escher, and B. Kolev, *Local and global well-posedness of the fractional order EPDiff equation*, J. Differential Equations, 258 (2015), pp. 2010–2053.

[6] M. Bauer and M. Bruveris eds., *Math in the Cabin—Shape Analysis Workshop in Bad Gastein*, Technical report hal-01076953, University of Vienna, Vienna, 2014.

[7] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*, Int. J. Comp. Vision, 61 (2005), pp. 139–157.

[8] M. Bruveris and F. Vialard, *On Completeness of Groups of Diffeomorphisms*, preprint, arXiv:1403.2089, 2014.

[9] C. J. BUDD, W. HUANG, AND R. D. RUSSELL, *Adaptivity with moving grids*, Acta Numer., 18 (2009), pp. 111–241.

[10] A. DOMINITZ AND A. TANNENBAUM, *Texture mapping via optimal mass transport*, IEEE Trans. Visualization Comp. Graph., 16 (2010), pp. 419–433.

[11] G. DZIUK AND C. M. ELLIOTT, *Finite element methods for surface PDEs*, Acta Numer., 22 (2013), pp. 289–396.

[12] D. G. EBIN AND J. E. MARSDEN, *Groups of diffeomorphisms and the notion of an incompressible fluid*, Ann. of Math. (2), 92 (1970), pp. 102–163.

[13] R. A. FISHER, *On the mathematical foundations of theoretical statistics*, R. Soc. Lond. Philos. Trans. Ser. A Math Phys. Eng. Sci., Containing Papers of a Mathematical or Physical Character, (1922), pp. 309–368.

[14] T. FRIEDRICH, *Die fisher-information und symplektische strukturen*, Math. Nachr., 153 (1991), pp. 273–296.

[15] A. L. GIBBS AND F. E. SU, *On choosing and bounding probability metrics*, Internat. Statist. Rev., 70 (2002), pp. 419–435.

[16] V. GORBUNOVA, J. SPORRING, P. LO, M. LOEVE, H. A. TIDDENS, M. NIELSEN, A. DIRKSEN, AND M. DE BRUIJNE, *Mass preserving image registration for lung {CT}*, Med. Image Anal., 16 (2012), pp. 786–795.

[17] S. HAKER, L. ZHU, A. TANNENBAUM, AND S. ANGENENT, *Optimal mass transport for registration and warping*, Int. J. Comp. Vis., 60 (2004), pp. 225–240.

[18] R. S. HAMILTON, *The inverse function theorem of Nash and Moser*, Bull. Amer. Math. Soc. (N.S.), 7 (1982), pp. 65–222.

[19] D. D. HOLM AND J. E. MARSDEN, *Momentum maps and measure-valued solutions (peakons, filaments, and sheets) for the EPDiff equation*, in The Breadth of Symplectic and Poisson Geometry, Progr. Math. 232, Birkhäuser Boston, Boston, MA, 2005, pp. 203–235.

[20] D. D. HOLM, A. TROUVÉ, AND L. YOUNES, *The Euler-Poincaré theory of metamorphosis*, Quart. Appl. Math., 67 (2009), pp. 661–685.

[21] S. JOSHI AND M. MILLER, *Landmark matching via large deformation diffeomorphisms*, IEEE Trans. Image Process., 9 (2000), pp. 1357–1370.

[22] B. KHESIN, J. LENELLS, G. MISIOŁEK, AND S. C. PRESTON, *Geometry of diffeomorphism groups, complete integrability and geometric statistics*, Geom. Funct. Anal., 23 (2013), pp. 334–366.

[23] B. KHESIN AND R. WENDT, *The Geometry of Infinite-Dimensional Groups*, Ergeb. Math. Grenzgeb. (3) 51, Springer-Verlag, Berlin, 2009.

[24] S. MARSLAND, R. I. MCLACHLAN, K. MODIN, AND M. PERLMUTTER, *On conformal variational problems and free boundary continua*, J. Phys. A, 47 (2014), 145204.

[25] P. MICHOR AND D. MUMFORD, *On Euler's equation and 'EPDiff'*, J. Geom. Mech., 5 (2013), pp. 319–344.

[26] M. I. MILLER AND L. YOUNES, *Group actions, homeomorphisms, and matching: A general framework*, Int. J. Comp. Vision, 41 (2001), pp. 61–84.

[27] K. MODIN, *Generalized Hunter–Saxton equations, optimal information transport, and factorization of diffeomorphisms*, J. Geom. Anal., 25 (2015), pp. 1306–1334.

[28] K. MODIN, M. PERLMUTTER, S. MARSLAND, AND R. I. MCLACHLAN, *On Euler-Arnold equations and totally geodesic subgroups*, J. Geom. Phys., 61 (2011), pp. 1446–1461.

[29] T. A. E. MOSELHY AND Y. M. MARZOUK, *Bayesian inference with optimal maps*, J Comput. Phys., 231 (2012), pp. 7815–7850.

[30] J. MOSER, *On the volume elements on a manifold*, Trans. Amer. Math. Soc., 120 (1965), pp. 286–294.

[31] F. OTTO, *The geometry of dissipative evolution equations: The porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[32] N. PAPADAKIS, G. PEYRÉ, AND E. OUDET, *Optimal transport with proximal splitting*, SIAM J. Imag. Sci., 7 (2014), pp. 212–238.

[33] J. RABIN, G. PEYRÉ, J. DELON, AND M. BERNOT, *Wasserstein barycenter and its application to texture mixing*, in Scale Space and Variational Methods in Computer Vision, Lecture Notes in Comput. Sci. 6667, Springer, Berlin, 2012, pp. 435–446.

[34] S. REICH, *A nonparametric ensemble transform method for Bayesian inference*, SIAM J. Sci. Comput., 35 (2013), pp. A2013–A2024.

[35] A. TROUVÉ AND L. YOUNES, *Metamorphoses through lie group action*, Found. Comput. Math., 5 (2005), pp. 173–198.

[36] A. TROUVÉ AND L. YOUNES, *Local geometry of deformable templates*, SIAM J. Math. Anal., 37 (2005), pp. 17–59.

[37] T. UR REHMAN, E. HABER, G. PRYOR, J. MELONAKOS, AND A. TANNENBAUM, *3d nonrigid registration via optimal mass transport on the GPU*, Medical Image Anal., 13 (2009), pp. 931–940.

[38] C. VILLANI, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer-Verlag, Berlin, 2009.

[39] Y. YIN, E. A. HOFFMAN, AND C.-L. LIN, *Mass preserving nonrigid registration of CT lung images using cubic B-spline*, Med. Phys., 36 (2009), pp. 4213–4222.

[40] L. YOUNES, *Shapes and Diffeomorphisms*, Appl. Math. Sci. 171, Springer-Verlag, Berlin, 2010.

[41] X. ZHAO, Z. SU, X. D. GU, A. KAUFMAN, J. SUN, J. GAO, AND F. LUO, *Area-preservation mapping using optimal mass transport*, IEEE Trans. Visualization Comp. Graph., 19 (2013), pp. 2838–2847.

[42] L. ZHU, Y. YANG, S. HAKER, AND A. TANNENBAUM, *An image morphing technique based on optimal mass preserving mapping*, IEEE Trans. image Process., 16 (2007), pp. 1481–1495.