Journal of Computational Physics 500 (2024) 112726

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Adaptive mesh methods on compact manifolds via Optimal Transport and Optimal Information Transport

Axel G.R. Turnquist

Department of Mathematics, University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Keywords:
Optimal transport
Optimal information transport
Diffeomorphic density matching
Moving mesh methods
Convergent numerical methods
Compact manifolds

ABSTRACT

Moving mesh methods were devised to redistribute a mesh in a smooth way, while keeping the number of vertices of the mesh and their connectivity unchanged. A fruitful theoretical point-of-view is to take such moving mesh methods and think of them as an application of the diffeomorphic density matching problem. Given two probability measures μ_0 and μ_1, the diffeomorphic density matching problem consists of finding a diffeomorphic pushforward map T such that $T_{\#}\mu_0 = \mu_1$. Moving mesh methods are seen to be an instance of the diffeomorphic density matching problem by treating the probability density as the local density of nodes in the mesh. It is preferable that the restructuring of the mesh be done in a smooth way that avoids tangling the connections between nodes, which would lead to numerical instability when the mesh is used in computational applications. This then suggests that a diffeomorphic map T is desirable to avoid tangling. The first tool employed to solve the moving mesh problem between source and target probability densities on the sphere was Optimal Transport (OT). Recently Optimal Information Transport (OIT) was rigorously derived and developed allowing for the computation of a diffeomorphic mapping by simply solving a Poisson equation. Not only is the equation simpler to solve numerically in OIT, but with Optimal Transport there is no guarantee that the mapping between probability density functions defines a diffeomorphism for general 2D compact manifolds.

In this manuscript, we perform a side-by-side comparison of using Optimal Transport and Optimal Information Transport on the sphere for adaptive mesh problems. We choose to perform this comparison with recently developed provably convergent solvers, but these are, of course, not the only numerical methods that may be used. We believe that Optimal Information Transport is preferable in computations due to the fact that the partial differential equation (PDE) solve step is simply a Poisson equation. For more general surfaces M, we show how the Optimal Transport and Optimal Information Transport problems can be reduced to solving on the sphere, provided that there exists a diffeomorphic mapping $\Phi : M \to \mathbb{S}^2$. This implies that the Optimal Transport problem on M with a special cost function can be solved with regularity guarantees, while computations for the problem are performed on the unit sphere.

In this manuscript, importantly, we think of moving mesh methods as a particular case of the diffeomorphic density matching problem. The diffeomorphic density matching problem has a long history in the imaging sciences. Classical methods consist of positive scalar image functions composed from the right by transformations [39]. That is, given a source probability density function f_0 and

E-mail address: agt6@njit.edu.

https://doi.org/10.1016/j.jcp.2023.112726
Received 22 December 2021; Received in revised form 19 December 2023; Accepted 19 December 2023

Available online 27 December 2023

a target probability density function f_1, classical techniques compute $f_0 \circ T = f_1$. Non-classical methods, like Optimal Transport and Optimal Information Transport, allow the transformation to act as a pushforward or pullback on the density, that is $f_0 = |DT| f_1 \circ T$ or $|DT| f_0 \circ T = f_1$, respectively. This generalization has particular benefit in that it allows for proving the existence of a diffeomorphic mapping over a perhaps surprisingly wide range of densities f_0 and f_1.

Non-classical diffeomorphic density matching has found widespread applications in medical image registration, see, for example, [13,14,8,32], random sampling from a probability measure see, for example, [3,27] and adaptive moving mesh methods, for example, in meteorology see the line of work in [4,38,5], among other applications. These applications are necessarily more challenging from a theoretical perspective as well as a computational perspective when the geometry is non-Euclidean. In this manuscript, we perform a comparison of Optimal Transport versus Optimal Information Transport for the moving mesh problem in the non-Euclidean case and provide a reformulation of the two problems, which are useful in cases beyond the sphere.

Optimal Transport density matching in non-Euclidean geometries is a field of study in which much progress on regularity theory has been made recently, see the following for examples of this large line of work [23,19,20,11,34,21]. Recently, the authors in [15, 16] introduced a numerical convergence framework and convergent numerical scheme for the PDE formulation of the optimal transport problem on the sphere which included, as one application, the traditional squared geodesic cost $c(x, y) = \frac{1}{2} d_{\mathbb{S}^2}(x, y)^2$. The convergence framework addressed situations where the solutions u of the PDE formulation are known to be *a priori* smooth (that is, at least C^2) as well as nonsmooth (but still C^1). This convergence framework required that the numerical schemes were consistent, stable in the average value, and monotone and achieved compactness via a Lipschitz constraint imposed on the PDE level. Preliminary results in [16] showed the success of using Optimal Transport on the sphere for moving mesh methods. In this manuscript, we demonstrate the results in cases where the provably convergent numerical scheme for Optimal Transport on the sphere in [16] has success and where possible problems arise.

Optimal Information Transport was developed in the papers [2,26]. This allows for a diffeomorphic map T between two smooth densities to be computed via solving a Poisson equation. Demonstrations of the success of this method are shown [2] for subsets of Euclidean space. Optimal Information Transport methods have subsequently found applications in image registration for medical applications, for example in [8,32] and also random sampling [3], also for subsets of Euclidean space. In this manuscript, we use a provably convergent scheme for the Optimal Information Transport problem on the sphere and show its success for the moving mesh problem in challenging cases on the unit sphere and on an oblate sphere.

In this manuscript, we utilize a class of provably convergent methods (with convergence rate guarantees) for computing the Optimal Information Transport on smooth, compact, and connected 2D manifolds M. We then provide a comparison of the adaptive mesh methods on these manifolds via Optimal Transport and Optimal Information Transport. We show how to reformulate the Optimal Transport and Optimal Information Transport problems on manifolds M that are diffeomorphic to the sphere, so as to solve the diffeomorphic density matching problem. In Section 1 we introduce the applied problem of adaptive moving mesh methods and show how the problem of finding a pushforward map T can be solved via Optimal Transport and Optimal Information Transport. In Section 2 we review briefly the literature on the problem of Optimal Transport on the sphere (and beyond) and highlight the lack of regularity guarantees for the Optimal Transport mapping T that arise by using the squared geodesic cost function. In Section 3 we outline how the Optimal Information Transport problem is derived in the literature. In Section 4, we introduce how, given a diffeomorphism $\Phi : M \to M_0$, our diffeomorphic density matching problem can be solved with Optimal Transport or Optimal Information Transport on M_0 and prove that this is equivalent to solving an Optimal Transport and Optimal Information Transport problem on M with a special cost function. In the case of genus-0 compact surfaces, this choice of cost function allows for smooth Optimal Transport mappings T as opposed to the squared geodesic cost function which has no such regularity guarantees. In Section 5, we show the grid setup and how consistent, monotone, and stable discrete operators are constructed for the PDE in Optimal Transport and Optimal Information Transport on the sphere. In Section 6 we demonstrate the following computations on the unit sphere: (1) a simple example of redistributing node density smoothly around the equator, (2) a difficult example of redistributing node density with a discontinuous globe density, (3) solving a problem for which the exact solution is known for the Optimal Transport problem, (4) an example where there is significant compression of the mesh, and (5) redistributing node density to be concentrated about the "north pole" of an ellipsoid. And in Section 7, we summarize the key contributions and results and explain future directions of investigation.

1. Moving mesh methods

Moving mesh methods are a type of adaptive mesh methods that have found special usage in computational PDE techniques. The transformed mesh is then often used in high-resolution PDE computations [4]. Fixing the number of grid points leads to simpler data structures (than some other adaptive mesh techniques) in the PDE solving step, as was stressed in the papers [4,38,7].

In the Euclidean setup, the moving mesh problem setup requires the "physical" target domain $\Omega_p \subset \mathbb{R}^d$, where the PDE is posed, while the "computational" input domain $\Omega_c \subset \mathbb{R}^d$ is usually chosen to be a uniform rectangular grid (in Euclidean space). In non-Euclidean geometries, one would like to use any simple mesh generator for the computational mesh. Then, the objective is to redistribute the points in the mesh to a desired density while avoiding the "tangling" of the mesh. Therefore, our task is to find a diffeomorphic mapping $T : \Omega_c \to \Omega_p$. Typically, in applications, the local density of grid points in the physical domain Ω_p is determined by, for example, the time scales involved in the solution of a fluid mechanics PDE. For example, in an evolving front or shock, it is desirable for the density of points in Ω_p to be greater than in the relatively unchanging parts of the solution, as in meteorological applications where it is desirable to have high resolution in areas of high precipitation [38]. This process could be done iteratively to get a very accurate solution of the PDE, by solving the shock on more and more resolved grids as one iterates. The

way this is done is by feeding the information from the PDE (desired density) into a scalar monitor function $\mathcal{M}(y,t) > 0$ and then solving the change of variables formula:

$$\mathcal{M}(y,t)J(T(x)) = \theta(t), \tag{1}$$

where J is the Jacobian of the diffeomorphic mapping T and $\int_{\Omega_p} \mathcal{M}(y,t)dy = \theta(t)$. Moving mesh methods require that the mapping be a diffeomorphism, which means that the Jacobian of the mapping T satisfies $0 < J(T) < \infty$. In other words, we want a smooth, invertible mapping T. This will prevent the mesh from tangling. An example of a monitor function constructed from information from a function $f(y,t)$ is the scaled arc-length function:

$$\mathcal{M}(y,t) = \sqrt{1 + \left|S\nabla_y f(y,t)\right|^2}, \tag{2}$$

where S is a normalization factor, see [4].

The moving mesh method can be thought of as a particular case of the diffeomorphic density matching problem, which is solved via the Jacobian equation. Let M be an n-dimensional Riemannian manifold. Let μ_0 and μ_1 be two probability measures on M with density functions f_0 and f_1, respectively. Under very general conditions, a weak (distributional) solution of the Jacobian equation (3), exists, see [35].

$$f_0(x) = f_1(T(x))J(T(x)). \tag{3}$$

While the existence theorems for solutions to the Jacobian equation stipulate the conditions for which we have a distributional solution of the monitor equation (1), it does not state that such a mapping T is unique or when it is diffeomorphic. It does show that, however, for fixed source and target mass distributions, there can be many such T. What we will find is that under certain assumptions on the source and target masses and manifolds, we can find diffeomorphic mappings using the techniques of Optimal Transport and Optimal Information Transport, which both satisfy the Jacobian equation and give us diffeomorphic mappings, appropriate for our moving mesh methods.

Now that we know that our objective is to solve the Jacobian equation and find a particular T that is a diffeomorphism, at this point the problem turns into *how* to find such a T. We need to address the following issues: 1.) Under what conditions should a diffeomorphism exist and 2.) which equation needs to be solved to find such a diffeomorphism. The regularity results in the literature are summarized in this section, but more details are presented in Section 2 and Section 3. For the reader that wishes to quickly understand what types of equations need to be solved (and defer the technical details to another time) we refer to the equations that need to be solved. The Optimal Transport mapping is computed by solving Equation (8), which is a fully nonlinear elliptic PDE for a potential function u and Equation (6) to relate the function u and the mapping T. The Optimal Information Transport mapping is computed using Equation (28), which has one PDE computation for a Poisson equation and then a first-order linear ODE which computes a path of diffeomorphisms $\varphi(t)$ that is a horizontal geodesic in the space of volume-preserving diffeomorphisms. Then the mapping T is found by finding the inverse time-1 map $T(x) = \varphi^{-1}(1)(x)$. The reader more interested in implementation and results can read this section for the summary of when a diffeomorphism exists, refer to the equations just mentioned and move directly to Section 4.

In the Euclidean case, the mapping arising from Optimal Transport is usually chosen by further requiring that the map T minimize the following integral:

$$I = \int_{\Omega_c} |T(x) - x|^2 \, d\mu_0(x). \tag{4}$$

It can be shown that such a mapping is the gradient of a convex function u, that is: $T(x) = \nabla_x u(x)$, which makes it irrotational (meaning $\nabla \times T = 0$ [5]) if the underlying manifold is $\mathbb{R}^d$. More specifically, the regularity theory of Caffarelli (see the summary in [33]) shows us that if $f_0, f_1 \in C^{0,\alpha}(\mathbb{R}^d)$ for $0 < \alpha < 1$ and $0 < f_0, f_1 < \infty$, then we are guaranteed that the mapping is, in fact, a diffeomorphism.

In the more general manifold setting, we have that the mapping solving the Optimal Transport problem with squared geodesic cost is given by $T(x) = \exp_x(\nabla u)$ (as long as the manifold is geodesically complete), see [23], where u is as a c-convex function, a technical definition that is common in the Optimal Transport literature, see [19] for a discussion of this condition and more about the general case of manifolds. For the n-sphere, fairly general conditions on the source and target masses are given in [20], which show that mass transports to a distance bounded strictly below the injectivity radius and gives us differentiability. In order for the mapping to be a diffeomorphism also need $0 < f_0, f_1 < \infty$ and $f_0, f_1 \in C^{1,1}(\mathbb{S}^2)$. More detail on the regularity results of the Optimal Transport problem will be shown in 2. The more general manifold with the squared geodesic cost function fails in terms of differentiability, see the result in [11] for a simple example of an oblate sphere where differentiability fails to be guaranteed.

Optimal Information Transport provides another option for computing a mapping that solves the diffeomorphic density matching problem. We will provide much more detail on the Optimal Information Transport problem in Section 3, but we summarize the relevant details here. In this case, given a geodesic $\mu(t)$ in the space of probability measures endowed with the infinite-dimensional Fisher-Rao metric, such that $\mu(0) = \mu_0$ and $\mu(1) = \mu_1$, where μ_0 and μ_1 are given positive and smooth probability measures, there exists a unique path of diffeomorphisms φ_t such that $\varphi_*(1)\mu_0 = \mu_1$. The key here is the manifold M can be any compact, connected and complete manifold and the equation for the path of diffeomorphisms $\varphi(t)$ is particularly simple (28).

For the moving mesh problem, then, we expect both Optimal Transport and Optimal Information Transport to be versatile and useful in $\mathbb{R}^d$ and $\mathbb{S}^d$. Beyond these cases, Optimal Information Transport is preferable. In Section 4, we offer means of reformulating these problems on more general manifolds M so that we have a hope of finding a diffeomorphic mapping.

Looking at moving mesh methods as a particular type of adaptive mesh methods, we can see that a possible disadvantage lies in the inability to add points in the mesh to refine the resolution. In a grid built from placing vertices at points of equally spaced longitude and latitude, for example, one cannot increase the number of vertices exactly along the equator without "tangling" the mesh. Thinking about moving mesh methods as diffeomorphic density matching between two C^∞ probability measures bounded away from zero, one can only "borrow" vertices from elsewhere in the mesh and create a higher local *density* of points along the equator. On the bright side, however, by using OT and OIT for moving mesh methods, the only really important quantity is the *ratio* of the source and target density. That is, starting with any given mesh, we can redistribute the density of vertices *as if the original mesh had constant vertex density*. Thus, the tradeoff in not being able to exactly specify where the new nodes go allows for tremendous flexibility in the local density of nodes, and in computations vertices can always be added to the grid, before or after the mesh redistribution, in order to refine the mesh as desired.

In applications, there is interest in applying moving mesh methods to the (non-compact) ocean. Suppose one begins with a mesh that is only over the oceans. Can one use diffeomorphic density matching techniques, such as Optimal Information Transport, in order to restructure this mesh over the ocean? Geometrically, Optimal Information Transport is posed over manifolds. Therefore, the natural choices are to pose the mesh on $M = \mathbb{S}^2$, or, perhaps, conformally map the oceans to a subset of $\mathbb{R}^2$ and then use the geometry from the torus $M = \mathbb{T}^2$. The source density in Optimal Information Transport is constant. The target density must be bounded away from zero in order to guarantee a diffeomorphic mapping T. This would lead to some mass-transfer across the boundary of the oceans. That is, some nodes leaving the oceans and entering the continents are inevitable (and therefore the number of nodes over the oceans will not be conserved). Using Optimal Transport it is also not possible to guarantee a diffeomorphic mapping for the mesh over the ocean. In this case, also, the source and target mass densities must be bounded away from zero, which will lead to the target mesh having some nodes, inescapably, over the continents. It is not possible either to map the oceans conformally, say, to a subset of Euclidean space, use source and target densities that have support on this set, and then solve the Monge-Ampère equation to guarantee the mapping T is diffeomorphic. This is because the set of oceans contains holes (continents) and is therefore not convex. Since the target set is not convex, the classical regularity results by Caffarelli in subsets of Euclidean space do not guarantee that the map is differentiable.

2. Optimal Transport on compact manifolds

Our main objective in this section is to review the regularity theory for the Optimal Transport mapping between probability measures on compact and connected 2D manifolds, since our goal is to know when the Optimal Transport mapping is a diffeomorphism (i.e. sufficiently smooth and invertible). The regularity results can largely be divided into two régimes. These two régimes are 1.) positive C^∞ source and target probability density functions lead to a C^∞ Optimal Transport mapping T and 2.) positive C^∞ source and target probability density functions *do not necessarily* lead to a C^∞ Optimal Transport mapping T. The first case (smooth mapping) is possible for the squared geodesic cost function on the n-sphere [20]. Beyond the n-sphere, there exist compact and connected 2D manifolds where a smooth mapping is not possible [11] when using the squared geodesic cost function. This suggests that either it is not possible to guarantee the existence of smooth mappings T or perhaps it is possible with a different cost function. For genus-0 manifolds, in Section 4, in fact, we make the assertion that a change of cost function will, in fact, guarantee regularity for the mapping T. The case beyond genus-0 manifolds will be explored in future work.

Since our objective is to study the mapping between probability distributions, we consider the Monge problem of Optimal Transport on smooth compact, connected manifolds M without boundary. The Monge problem of Optimal Transport considers finding the mapping which minimizes the total cost of moving mass from a given source probability measure to a target probability measure. This total cost is computed by integrating a cost function over the source measure. Optimal Transport allows for a wide variety of cost functions to be chosen. In this manuscript we will often discuss the squared geodesic cost, that is $c(x,y) = \frac{1}{2}d_M(x,y)$. However, it should be strongly emphasized here that *any* cost function which allows one to construct a diffeomorphic pushforward map $T_\# \mu_0 = \mu_1$ is possible to use. For the unit sphere $\mathbb{S}^2$ this includes, but is not limited to, any cost function that satisfies the hypotheses of Theorem 4.1 from [20], as well as the reflector antenna cost $c(x,y) = -\log\|x-y\|$, where $\|\cdot\|$ is the vector 2-norm in $\mathbb{R}^3$, again, see [20] for more details.

Given source and target probability measures μ_0 and μ_1 supported on the manifold that have density functions f_0 and f_1, respectively, that is $d\mu_0(x) = f_0(x)\text{dvol}(x)$ and $d\mu_1(x) = f_1(y)\text{dvol}(y)$, the Monge Problem of Optimal Transport consists of finding a mapping $T : M \to M$ such that $T_\# \mu_0 = \mu_1$, that is μ_1 is the pushforward measure of μ_0 via the mapping T such that:

$$T = \underset{S}{\text{argmin}} \int_M \frac{1}{2} d_M(x, S(x))\, f_0(x) dx. \tag{5}$$

The Monge problem of Optimal Transport can be formulated as a partial differential equation given some additional assumptions on the cost function, known as the MTW conditions, which are satisfied by the squared geodesic cost function [22].

Hypothesis 1 *(Conditions on cost function).* The squared geodesic cost function $c(x,y) = \frac{1}{2}d_M(x,y)$ satisfies the following conditions:

(a) The function $\nabla_{M,x} c(x,y)$ is Lipschitz for all $x, y \in M$ and invertible in y,

(b) For any $x, y \in M$, we have $\det D^2_{M,xy} c(x,y) \neq 0$.

As first shown in [23], the optimal map corresponding to this cost function is determined from the condition

$$T(x) = \exp_x(\nabla_M u(x)), \ x \in M, \tag{6}$$

which, for the unit sphere $\mathbb{S}^2$ has the explicit expression:

$$T(x) = x \cos\left(\left\|\nabla_{\mathbb{S}^2} u(x)\right\|\right) + \frac{\nabla_{\mathbb{S}^2} u(x)}{\left\|\nabla_{\mathbb{S}^2} u(x)\right\|} \sin\left(\left\|\nabla_{\mathbb{S}^2} u(x)\right\|\right) \tag{7}$$

where u is the solution of the fully nonlinear partial differential equation

$$F\left(x, \nabla_M u(x), D^2_M u(x)\right) = 0, \tag{8}$$

where

$$F(x,p,M) \equiv -\det\left(M + A(x,p)\right) + H(x,p), \tag{9}$$

and

$$A(x,p) = D^2_{M,xx} c\left(x, T(x,p)\right)$$
$$H(x,p) = \left|\det D^2_{M,xy} c\left(x, T(x,p)\right)\right| f_0(x) / f_1\left(T(x,p)\right).$$

The PDE (8) is also subject to the c-convexity condition, which requires that

$$D^2_M u(x) + A(x, \nabla_M u(x)) \geq 0. \tag{10}$$

2.1. Regularity

We consider the optimal transport problem (8) under the following hypothesis.

Hypothesis 2 *(Conditions on data (smooth))*. We require problem data to satisfy the following conditions:

(a) There exists some $m > 0$ such that $f_0(x) \geq m$ and $f_1(x) \geq m$ for all $x \in \mathbb{S}^n$.
(b) The mass balance condition holds, $\int_{\mathbb{S}^2} f_0(x)\, dx = \int_{\mathbb{S}^2} f_1(y)\, dy$.
(c) The cost function is the squared geodesic cost function $c(x,y) = \frac{1}{2} d_{\mathbb{S}^2}(x,y)^2$.
(d) The data satisfies the regularity requirements $f_0, f_1 \in C^{1,1}(\mathbb{S}^n)$.
(e) The geometry is the n-sphere $\mathbb{S}^n$.

We thus get the following regularity result, from Loeper [20]:

Theorem 3 *(Regularity). The optimal transport problem* (8) *with data satisfying Hypothesis* 2 *has a classical solution* $u \in C^3(\mathbb{S}^2)$. *Furthermore, if* $f_0, f_1 \in C^\infty(\mathbb{S}^2)$ *and positive, then* $u \in C^\infty(\mathbb{S}^2)$.

Such Optimal Mapping T is also a diffeomorphism. This is because T is measure preserving, which means it solves the Jacobian equation:

$$f_0(x) = f_1\left(T(x)\right) |DT|, \tag{11}$$

which implies:

$$|DT| = f_0(x) / f_1(T(x)) > 0, \tag{12}$$

by the assumption that f_0 and f_1 are bounded and positive given in Hypothesis 2.

In Section 6, we will perform a computation cases where the source and target densities are piecewise continuous. So, for our purposes, the weakest regularity guarantees for the Optimal Transport problem on the sphere are that T is merely Hölder continuous, see Theorem 2.4 in Loeper [20] for more details.

The solution to (8) is unique only up to additive constants. In this manuscript, for computations, we will fix a point $x_0 \in M$ and add the additional constraint:

$$u(x_0) = 0. \tag{13}$$

2.2. Beyond the sphere

The regularity guarantees may fail once we move beyond the n-sphere. There exist compact and connected 2D manifolds for which the same regularity results apply as long as they are small perturbations of the sphere [34], however in general even with positive C^∞ source and target masses, we cannot be guaranteed a smooth mapping T with even the squared geodesic cost function. The mapping T may fail to be smooth when the so-called cost-sectional curvature is found to be negative for some point $x \in M$. In fact, in some cases, manifolds M with non-positive cost-sectional curvature at any point $x \in M$ can have positive measures $\mu_0, \mu_1 \in C^\infty(M)$, but T is not even guaranteed to be continuous when the cost function is the squared geodesic cost [19].

In order to define cost-sectional curvature, we introduce the Ma, Trudinger, Wang tensor [22,19]:

$$\mathfrak{G}_c(x_0, y_0)(\xi, \nu) = D^4_{p_\nu, p_\nu, x_\xi, x_\xi}\left[(x,p) \mapsto -c\left(x, T_{x_0}(p)\right)\right]|_{x_0, p_0 = -\nabla_x x(x_0, y_0)}. \tag{14}$$

The cost-sectional curvature is negative at a point (x, y) if there exist ξ, ν such that $\mathfrak{G}_c(x,y)(\xi,\nu) < C_0 |\xi|^2 |\nu|^2$. In this case of the squared geodesic cost function, the non-positivity of the cost-sectional curvature is equivalent to the negativity of the sectional curvature of M at any point $x \in M$ due to the equality:

$$\frac{\mathfrak{G}_c(x,x)(\nu,\xi)}{|\xi|^2 |\nu|^2 - (\xi \cdot \nu)^2} = \frac{2}{3} \cdot \text{sectional curvature of } M \text{ at } x \text{ in the plane } (\xi, \nu). \tag{15}$$

Notice in (15) that we are looking at the diagonal (where $x = y$), but even if the underlying manifold has strictly positive curvature everywhere, the Ma, Trudinger, Wang tensor $\mathfrak{G}_c(x,y)$ may be negative on the off-diagonal (x,y) s.t $x \neq y$. This was shown true even for some ellipsoids of revolution in the paper [11]. This suggests that computing Optimal Transport for such cases using the squared geodesic cost function is not guaranteed to result in a diffeomorphic mapping T.

3. Optimal Information Transport on compact manifolds

Here we review the Optimal Information Transport problem on a compact and connected manifold M, which yields a diffeomorphic mapping as long as the source and target masses are positive and smooth. The benefit here is that the diffeomorphism exists over all such M, contrasting with the Optimal Transport mapping. For much more information about Optimal Information Transport, see the background and more detail presented in [2,26]. For all the discussion in this section, we will assume that the underlying manifold M is compact and connected and is equipped with the standard volume form vol.

3.1. Geometry of space of diffeomorphisms on M and space of positive smooth densities on M

In order to find a diffeomorphism between two positive, smooth probability measures, one approach is to work directly in the space of diffeomorphisms. Denote by $\text{Diff}(M)$ the set of diffeomorphisms from the manifold M to itself. This set is nonempty, since $\text{Id} \in \text{Diff}(M)$. Also, if $\varphi_1, \varphi_2 \in \text{Diff}(M)$, then $\varphi_1 \circ \varphi_2 \in \text{Diff}(M)$. Also, for every $\varphi \in \text{Diff}(M)$, there exists a $\varphi^{-1} \in \text{Diff}(M)$ such that $\varphi \circ \varphi^{-1} = \text{Id}$. Thus, $\text{Diff}(M)$ has an algebraic structure and, equipped with a metric g, is a Lie group. Denote by $\mathfrak{X}(M)$ the space of smooth vector fields on M with the "information" inner product $\langle U, V \rangle^I_\varphi$ for a fixed $\varphi \in \text{Diff}(M)$ and $U, V \in \mathcal{T}_\varphi$:

$$\langle U, V \rangle^I_\varphi = \int_M g(\Delta u, v) dx + \lambda \sum_{i=1}^k \int_M g(u, \xi_i) dx \int_M g(v, \xi_i) dx, \tag{16}$$

where g is the induced metric on M, $u = U \circ \varphi^{-1}, v = V \circ \varphi^{-1}$, $\lambda > 0$, $u = U \circ \varphi^{-1}$, $v = V \circ \varphi^{-1}$, g is the underlying metric of the manifold, Δ is the Laplace-de Rham operator (Hodge Laplacian) on the space of vector fields, and $\{\xi_i\}_i$ is an orthonormal basis of the harmonic vector fields on M, that is those vector fields ξ for which $\Delta \xi = 0$.

Notice there is a natural "connection" between $\text{Diff}(M)$ and measures defined on M. That is, fix a positive, smooth probability density on the manifold μ, then there exist measure-preserving mappings (for example, the identity map Id) φ_μ such that $\varphi^*_\mu \mu = \mu$. That is,

$$\int_E \mu(x) dx = \int_{\varphi_\mu(E)} \mu(y) dy, \tag{17}$$

for any Borel set $E \subset M$. For a fixed μ, these measure-preserving diffeomorphisms φ_μ form a subgroup of $\text{Diff}(M)$.

Now, we define the space of smooth volume forms over M with total volume $\text{vol}(M)$:

$$\text{Dens}(M) = \left\{ \mu \in \Omega^n(M) : \int_M \mu = \text{vol}(M), \mu > 0 \right\}, \tag{18}$$

where the notation $\Omega^n(M)$ denotes the n-forms over the manifold M. The metric we choose on this space is the Fisher-Rao metric, defined by:

$$\langle \sigma_1, \sigma_2 \rangle_\mu^F = \frac{1}{4} \int_M \frac{d\sigma_1}{d\mu} \frac{d\sigma_2}{d\mu} d\mu, \tag{19}$$

where $\frac{d\sigma_i}{d\mu}$ denotes the Radon-Nikodym derivative of σ_i with respect to μ. The space $\mathrm{Dens}(M)$ is a Fréchet Manifold.

With this choice of metric, for compact and connected manifolds M, the geodesics in $\mathrm{Dens}(M)$ have explicit expressions. Define $W : \mathrm{Dens}(M) \to C^\infty(M)$ by $\mu \mapsto \sqrt{\frac{d\mu}{d\mathrm{vol}}}$, then the geodesic $\mu(t)$ between $\mu_0, \mu_1 \in \mathrm{Dens}(M)$ are given by the mapping:

$$[0,1] \ni t \mapsto \left(\frac{\sin((1-t)\theta)}{\sin\theta} f_0 + \frac{\sin(t\theta)}{\sin\theta} f_1 \right)^2 \mathrm{vol}, \tag{20}$$

where

$$\theta = \arccos\left(\frac{\langle f_0, f_1 \rangle_{L^2}}{\mathrm{vol}(M)} \right), \tag{21}$$

and $f_i = W(\mu_i)$.

The question then can be asked if there are geodesics in $\mathrm{Diff}(M)$ that descend onto the explicit geodesic (20), therefore yielding a path $\varphi(t) \subset \mathrm{Diff}(M)$ such that $\varphi(t)^* \mu_0 = \mu(t)$. The answer is yes (for the rather complicated looking metric (16)), and, when $\mu_0 = \mathrm{vol}$, the expression to find $\varphi(t)$ is particularly simple!

Fix a reference measure $\mu \in \mathrm{Dens}(M)$. Then, define the space $\mathrm{Diff}_\mu(M)$ consisting of μ-preserving mappings on M. As noted before, this is a subspace of $\mathrm{Diff}(M)$, so we can define the quotient space $\mathrm{Diff}_\mu(M) \setminus \mathrm{Diff}(M)$ that consists of elements of the form:

$$[\varphi] := \mathrm{Diff}_\mu(M) \circ \varphi \in \mathrm{Diff}_\mu(M) \setminus \mathrm{Diff}(M). \tag{22}$$

An important result from Moser says that the map $\mathrm{Diff}_\mu(M) \setminus \mathrm{Diff}(M) \ni [\varphi] \mapsto \varphi^* \mu \in \mathrm{Dens}(M)$ is bijective [28].

Again, μ is fixed. Define the projection:

$$\pi_\mu : \mathrm{Diff}(M) \ni \varphi \mapsto \varphi^* \mu \in \mathrm{Dens}(M). \tag{23}$$

Then we can define the following principal bundle structure:

$$\begin{array}{ccc} \mathrm{Diff}_\mu(M) & \hookrightarrow & \mathrm{Diff}(M) \\ & & \downarrow \pi_\mu \\ & & \mathrm{Dens}(M) \end{array}$$

Now, fix $\mu = \mathrm{vol}$, then we denote the projection $\pi_{\mathrm{vol}} = \pi$ and we will observe the important properties that result. We define the vertical distribution $\mathcal{V}_\varphi$, which is a subbundle of the tangent bundle $\mathcal{T}_\varphi$ that is tangent to the fibers. The horizontal distribution $\mathcal{H}_\varphi$ consists of all $U \in \mathcal{T}_\varphi$ such that $G_\varphi^I(U, V) = 0$, where V is in the vertical distribution. The horizontal distribution has a special characterization:

$$\mathcal{H}_\varphi = \left\{ U \in \mathcal{T}_\varphi \mathrm{Diff}(M) : U \circ \varphi^{-1} = \mathrm{grad}(f), f \in C^\infty(M) \right\}, \tag{24}$$

and if $\varphi(t)$ is a geodesic (in $\mathrm{Diff}(M)$), then if $\dot\varphi(0) \in \mathcal{H}_{\varphi(0)}$, then $\dot\varphi(t) \in \mathcal{H}_{\varphi(t)}$. Most critically, if $\varphi(t)$ is a horizontal geodesic curve in $\mathrm{Diff}(M)$, then $\mu(t) = \pi(\varphi(t))$ is a geodesic curve in $\mathrm{Dens}(M)$.

3.2. Diffeomorphic density matching

At this point, we will utilize the geometric properties we noted in Section 3.1 to derive simple equations for finding a horizontal geodesic curve $\varphi(t)$ such that $\mu(t) = \pi(\varphi(t))$ is a geodesic curve in $\mathrm{Dens}(M)$. Specifically, given a geodesic curve $\mu(t)$, we desire to find a $\varphi(t)$ that descends to $\mu(t)$. To find a unique path $\varphi(t)$, we also require that the path minimize an energy functional:

$$\begin{aligned} &\varphi(0) = \mathrm{Id}, \\ &\varphi^* \mu_0 = \mu(t), \\ &\text{minimizing} \int_0^1 \langle \dot\varphi(t), \dot\varphi(t) \rangle_{\varphi(t)}^I \, dt. \end{aligned} \tag{25}$$

From the observations in Section 3.1, we note that when $\mu = \mathrm{vol}$ that the horizontal geodesics have a particularly simple form given by Equation (24). Thus, from now on, we assume that μ_0 is the canonical volume form. We take the equation $\varphi^*(t)\mu_0 = \varphi^*(t)\mathrm{vol} = \mu(t)$ and differentiate with respect to t:

$$\dot\mu(t) = \partial_t \left(\varphi(t)^* \mathrm{vol} \right) = \varphi^* \mathrm{div}_{\mathrm{vol}} v(t), \tag{26}$$

where $v(t) = \dot{\varphi} \circ \varphi^{-1}$. This can be rewritten as:

$$\dot{\mu}(t) = \operatorname{div}(v(t)) \circ \varphi(t) \mu(t). \tag{27}$$

From here we write a Hodge-Helmholz decomposition for the vector field v, by writing $v = \operatorname{grad} f + w$. We look for a horizontal geodesic, so $\dot{\varphi}(0) \in \mathcal{H}_{\mathrm{Id}}$ and therefore $\dot{\varphi}(t) \in \mathcal{H}_{\varphi(t)}$ as was noted in 3.1 and thus from Equation (24), we see that we can choose $w = 0$. Therefore, the expression in Equation (27) simplifies to Equation (28) solving for the curl-free term f, with all other terms known, see Equation (20) for the equation for $\mu(t)$ and $\dot{\mu}(t)$.

$$\begin{aligned} \Delta f(t) &= \frac{\dot{\mu}(t)}{\mu(t)} \circ \varphi(t)^{-1}, \\ \dot{\varphi}(t) &= \operatorname{grad}(f(t)) \circ \varphi(t), \quad \varphi(0) = \mathrm{id}. \end{aligned} \tag{28}$$

This system of equations (28) is essentially the equation for the horizontal geodesic $\varphi(t)$. The Optimal Information Transport problem is then to find a diffeomorphism T such that:

$$T_* \mu_0 = \mu_1, \tag{29}$$

where $T_* \mu_0$ denotes the pushforward of μ_0, which is then given by the time-1 map: $T = \varphi^{-1}(1)$, where $\varphi(t)$ solves Equation (28).

4. Simplifying the geometry for moving mesh methods

The numerical methods proposed below in Section 5 show that diffeomorphic density matching for moving mesh methods computations via Optimal Transport and Optimal Information Transport both rely on some computational analysis of the underlying manifold M. They require precise computations of geodesics of M and Optimal Information Transport computations rely heavily on a well-chosen interpolation on the manifold M. In order to simplify greatly the amount of geometric analysis that must be performed, here we have found it useful to simplify the geometry by assuming that we have a diffeomorphic mapping $\Phi : M \to M_0$, where M_0 is a manifold where computations are simpler to perform. Since the smooth, connected, and compact surfaces M we address in this manuscript lie in $\mathbb{R}^3$, any two such manifolds that are homeomorphic are automatically diffeomorphic, see [18]. The first example of two homeomorphic manifolds that are *not* diffeomorphic, for example, was found by Milnor in 1956, see [25], in $\mathbb{R}^7$. This means that for our purposes, any two such smooth surfaces M that we wish to perform computations on that are of the same homotopy class are also diffeomorphic. Therefore, provided that M is, say, genus-0, we have that there exists a diffeomorphism $\Phi : M \to \mathbb{S}^2$. Constructing the diffeomorphism is another matter, although we show a simple example below in Section 4.6.

For example, for all smooth, compact genus-0 surfaces, we have found it particularly useful to choose $M_0 = \mathbb{S}^2$. So instead of performing computations on each individual ellipsoid M, we perform computations on the unit sphere, taking advantage, especially, of the fact that the sphere has explicit formulas for the geodesics. In this section we outline how, given two probability measures $\mu_0, \mu_1 \subset M$ with density functions $\bar{f}_0, \bar{f}_1 : M \to \mathbb{R}$, respectively, we may produce a diffeomorphic mapping $T : M \to M$ such that $T_\# \mu_0 = \mu_1$, by instead performing computations on M_0, given a diffeomorphic mapping $\Phi : M \to M_0$. For the sake of exposition, in this section, we will only show the theory for $M_0 = \mathbb{S}^2$.

This also has implications for solving Optimal Transport on genus-0 manifolds M for the purposes of producing a smooth Optimal Transport mapping T. As was mentioned in Section 2, if one desires to solve the Optimal Transport problem on a manifold M with the squared geodesic cost (which is often the natural choice in computations), there is no guarantee that the mapping T will be smooth, again see [11] for an example of an ellipsoid. In this section, we will show that if one has a diffeomorphism $\Phi : M \to \mathbb{S}^2$, then, letting $z, \xi \in M$, solving the Optimal Transport problem with the cost function $c(z, \xi) = \frac{1}{2}(\arccos(\Phi(z) \cdot \Phi(\xi)))^2$, where $\cdot$ denotes the dot product in $\mathbb{R}^3$ will guarantee a smooth Optimal Transport mapping T when $\bar{f}_0, \bar{f}_1$ are sufficiently smooth.

4.1. Choice of density functions on $\mathbb{S}^2$

Let x denote the coordinates on $\mathbb{S}^2$ and z denote the coordinates on M. Therefore, in coordinates, $\Phi(z) = x$. The density functions $\bar{f}_0, \bar{f}_1$ satisfy $\int_M \bar{f}_0 dx = \int_M \bar{f}_1 dx = 1$. Let $d\Phi_z : T_z M \to T_x \mathbb{S}^2$ denote the (pushforward) differential of the map Φ and $d\Phi_x^{-1} : T_x \mathbb{S}^2 \to T_z M$ the differential of the inverse function Φ^{-1}. With a choice of coordinates, $d\Phi_z$ is simply the Jacobian matrix of Φ and $\left|d\Phi_z\right|$ denotes its determinant. Then, for the Optimal Transport problem, we can define the following density functions on $\mathbb{S}^2$:

$$\begin{aligned} f_0(x) &= \bar{f}_0(\Phi^{-1}(x)) \left|d\Phi_x^{-1}\right| \\ f_1(x) &= \bar{f}_1(\Phi^{-1}(x)) \left|d\Phi_x^{-1}\right|. \end{aligned} \tag{30}$$

Then,

$$\int_{\mathbb{S}^2} f_0(x) dx = \int_{\Phi^{-1}(\mathbb{S}^2)} f_0(\Phi(z)) \left|d\Phi_z\right| dz, \tag{31}$$

$$= \int_{\Phi^{-1}(\mathbb{S}^2)} \bar{f}_0(z) dz = \int_M \bar{f}_0(z) dz = 1, \tag{32}$$

and likewise $\int_{\mathbb{S}^2} f_1(x)dx = 1$. Therefore, f_0, f_1 are density functions on $\mathbb{S}^2$ that integrate to 1.

For Optimal Information Transport, we perform computations on the unit sphere $\mathbb{S}^2$ where we need for $\mu_0 = \text{vol}$ in order to use the equations (28). Therefore, although we would like to perform computations for densities on M, we start in the other direction by setting $f_0 = 1$. Then, we define:

$$\begin{aligned} \bar{f}_0(z) &= \frac{\text{vol}(M)}{4\pi} \left| d\Phi_z \right| \\ \bar{f}_1(z) &= \frac{\text{vol}(M)}{4\pi} \left| d\Phi_z \right| f_1(\Phi(z)). \end{aligned} \tag{33}$$

Then,

$$\int_M \bar{f}_0(z)dz = \frac{\text{vol}(M)}{4\pi} \int_M \left| d\Phi_z \right| dz = \frac{\text{vol}(M)}{4\pi} \int_{\mathbb{S}^2} dx = \text{vol}(M) \tag{34}$$

and

$$\int_M \bar{f}_1(z)dz = \frac{\text{vol}(M)}{4\pi} \int_M \left| d\Phi_z \right| f_1(\Phi(z))dz = \frac{\text{vol}(M)}{4\pi} \int_{\mathbb{S}^2} f_1(x)dx = \text{vol}(M) \tag{35}$$

4.2. Producing a mass-preserving mapping $\bar{T}$ on M

In this section, we show that if we have a mass-preserving diffeomorphic mapping T on $\mathbb{S}^2$, then we can use it to build a mass-preserving diffeomorphic mapping $\bar{T}$ on M. This result will apply to both Optimal Transport and Optimal Information Transport mappings.

Lemma 4. *Given that T is mass preserving and diffeomorphic on $\mathbb{S}^2$ for f_0, f_1, the map $\bar{T}(z) = T(\Phi(z))$ is mass preserving and diffeomorphic on M for $\bar{f}_0, \bar{f}_1$.*

Proof. Since T is mass preserving, for any Borel set $B \subset M$, we have that:

$$\int_{\Phi(B)} f_0(x)dx = \int_{T\circ\Phi(B))} f_1(y)dy. \tag{36}$$

Therefore, for both the Optimal Transport and the Optimal Information Transport problems,

$$\int_{\Phi(B)} \bar{f}_0(\Phi^{-1}(x)) \left| d\Phi_x^{-1} \right| dx = \int_{T\circ\Phi(B)} \bar{f}_1(\Phi^{-1}(y)) \left| d\Phi_y^{-1} \right| dy. \tag{37}$$

Using the change of variables $\Phi(z) = x$ and $\Phi(\xi) = y$, we get:

$$\int_B \bar{f}_0(z)dz = \int_{\bar{T}(B)} \bar{f}_1(\xi)d\xi, \tag{38}$$

and therefore, $\bar{T}$ is mass-preserving for $\bar{f}_0, \bar{f}_1$ on M. □

If T is a diffeomorphism, then $\bar{T} = T\circ\Phi$ is a diffeomorphism. Therefore, since the moving mesh problem is simply concerned with changing the local density of the mesh, if we desire to change the local density of nodes by a factor f_1, we can solve the Optimal Information Transport problem using the densities $f_0 = 1$ and f_1 on the unit sphere $\mathbb{S}^2$ for a mapping T, then $T\circ\Phi(z)$ gives the pushforward mapping that locally changes the density by $f_1(\Phi(z))$ on the genus-0 manifold M. Beyond just thinking about T as simply a pushforward map, what is more interesting is that we can define a new Optimal Transport problem on M.

4.3. A good choice for the Optimal Transport cost function on M

We will show that by using a new cost function inherited from the unit sphere $\mathbb{S}^2$, we will be solving a new Optimal Transport problem on the manifold M. Here we outline a good choice for a class of cost functions $\bar{c} : M \times M \to \mathbb{R}$.

The squared geodesic cost function $c : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$ on the unit sphere is:

$$c(x, y) = \frac{1}{2} \left(\arccos(x \cdot y) \right)^2, \tag{39}$$

where $\cdot$ denotes the dot product in $\mathbb{R}^3$. We define a new cost function $\bar{c} : M \times M \to \mathbb{R}$ on the manifold M for points $z, \xi \in M$ by:

$$\bar{c}(z, \xi) = \frac{1}{2} \left(\arccos\left(\Phi(z) \cdot \Phi(\xi)\right) \right)^2 = c\left(\Phi(z), \Phi(\xi)\right). \tag{40}$$

It should be noted here that the choice of the diffeomorphism $\Phi : M \to \mathbb{S}^2$ is not necessarily unique, allowing for many such cost functions $\bar{c}$. Also, any cost function in Theorem 4.1 of [20], as well as the logarithmic cost appearing in the reflector antenna can be used as c.

4.4. Producing an optimal mapping $\bar{T}$ for the Optimal Transport problem on M

First, we define the two Optimal Transport problems, one on $\mathbb{S}^2$ and the other on M.
Define $\bar{\mu}_0, \bar{\mu}_1$ by $\frac{d\bar{\mu}_0}{d\text{vol}_M} = \bar{f}_0(x)$, and $\frac{d\bar{\mu}_1}{d\text{vol}_M} = \bar{f}_1(x)$. The mapping T solves the following Optimal Transport problem:

$$T = \text{argmin}_{S : S_{\#}\mu_0 = \mu_1} \int_{\mathbb{S}^2} c(x, S(x)) f_0(x) dx. \tag{OT'}$$

And there exists a $\tilde{T}$, see [23], solving the problem:

$$\tilde{T} = \text{argmin}_{S : S_{\#}\bar{\mu}_0 = \bar{\mu}_1} \int_{M} \bar{c}(x, S(x)) \bar{f}_0(x) dx. \tag{OT}$$

We desire to show that $\bar{T} = \tilde{T}$. We have already shown that $\bar{T}$ is mass-preserving. Define the function $\bar{u}$ as a function (up to a constant) that solves the following equation:

$$\nabla_M \bar{u}(z) = \nabla_{M,z} \bar{c}(z, \xi)|_{\xi = \bar{T}(z)}. \tag{41}$$

It remains to show that any $\bar{u}$ solving Equation (41) is $\bar{c}$-convex. Since Φ is a diffeomorphism, $d\Phi_z : T_z M \to T_x \mathbb{S}^2$ is invertible. Therefore, defining the function $\tilde{u}(z) = u(\Phi(z))$, we get:

$$\nabla_M \tilde{u}(z) = d\Phi_x^{-1} \circ \nabla_{\mathbb{S}^2} u(x)|_{x=\Phi(z)}, \tag{42}$$

$$= d\Phi_x^{-1} \circ \left(\nabla_{\mathbb{S}^2, x} c(x, y)|_{y = T(x)} \right) |_{x=\Phi(z)}, \tag{43}$$

$$= \nabla_{M,z} c\left(\Phi(z), \Phi(\xi)\right) |_{\xi = T \circ \Phi(z) = \bar{T}(z)}, \tag{44}$$

and therefore, $\nabla_M \tilde{u}(z) = \nabla_M \bar{u}(z)$. Therefore, $\bar{u} = u(\Phi(z)) + C$, for some arbitrary constant C.
We show that $\bar{u}$ is $\bar{c}$-convex.

Lemma 5. *Given that u is c-convex on $\mathbb{S}^2$, the function $\bar{u}$ is $\bar{c}$-convex on M.*

Proof. Since u is c-convex on $\mathbb{S}^2$, we know that for each $x \in \mathbb{S}^2$, there exists a $y \in \mathbb{S}^2$ such that:

$$\begin{cases} -u^c(y) - c(x, y) = u(x) \\ -u^c(y) - c(x', y) \le u(x'), \ \forall x' \in \mathbb{S}^2. \end{cases} \tag{45}$$

Fix $x \in \mathbb{S}^2$ and a diffeomorphism $\Phi : M \to \mathbb{S}^2$ and let $\Phi(z) = x$. Let $\xi \in M$ satisfy $\Phi(\xi) = y$. Then,

$$-\bar{u}^{\bar{c}}(\xi) - \bar{c}(z, \xi) = -\sup_{\tilde{z}} (-\bar{c}(\tilde{z}, \xi) - \bar{u}(\tilde{z})) - \bar{c}(z, \xi), \tag{46}$$

$$= -\sup_{\tilde{z}} (-c(\Phi(\tilde{z}), \Phi(\xi)) - u(\Phi(\tilde{z})) - C) - c(\Phi(z), \Phi(\xi)), \tag{47}$$

$$= -\sup_{\tilde{y}} (-c(\tilde{y}, y) - u(\tilde{y} - C)) - c(x, y), \tag{48}$$

$$= -u^c(y) + C - c(x, y) = u(x) + C = u(\Phi(z)) - C + C = \bar{u}(z). \tag{49}$$

Fix $x \in \mathbb{S}^2$ and a diffeomorphism $\Phi : M \to \mathbb{S}^2$ and let $\Phi(z) = x$. Let $\xi \in M$ satisfy $\Phi(\xi) = y$ and let x' be any point in $\mathbb{S}^2$, and let $\Phi(z') = x'$. Then,

$$-\bar{u}^{\bar{c}}(\xi) - \bar{c}(z', \xi) = -\sup_{\tilde{z}} (-\bar{c}(\tilde{z}, \xi) - \bar{u}(\tilde{z})) - \bar{c}(z', \xi), \tag{50}$$

$$= -\sup_{\tilde{z}} (-c(\Phi(\tilde{z}), \Phi(\xi)) - u(\Phi(\tilde{z})) - C) - c(\Phi(z'), \Phi(\xi)), \tag{51}$$

$$= -\sup_{\tilde{y}} (-c(\tilde{y}, y) - u(\tilde{y}) - C) - c(x', y), \tag{52}$$

$$= -u^c(y) + C - c(x', y) \le C + u(x'), \tag{53}$$

$$= C + u(\Phi(z')) = C + \bar{u}(z') - C = \bar{u}(z'). \quad \square \tag{54}$$

Therefore, $\bar{u}$ is $\bar{c}$-convex and thus $\bar{T} = T(\Phi(z))$ is a solution to the Optimal Transport problem on M with cost function $\bar{c}$.

Of course, (OT) and (OT') are different problems. If the application one has in mind requires anything but the mapping T, then solving (OT') as a surrogate for (OT) could certainly be inappropriate. However, for the purposes of mesh redistribution, it is not advisable to use the squared geodesic cost on M, as it is potentially fraught with regularity issues, whereas using $\bar{c}$ is not, since it yields an Optimal Transport map T which is a diffeomorphism by the guarantees from Theorem 3.

4.5. Diffeomorphic density matching for genus-1 compact surfaces

For genus-1 compact and connected surfaces (surfaces with one hole), like the torus, the computational scheme to solve the Optimal Transport problem or Optimal Information Transport problem just needs to be developed for a single M_0, then all computations can be done on a representative torus. For Optimal Transport on the torus, we do not have the same regularity guarantees that we did for the unit sphere. Therefore, we recommend to solve the diffeomorphic density matching problem via Optimal Information Transport in these cases, which will be explored in future work.

Exploring these kinds of computations is perhaps useful in real-world examples where data is known *a priori* to lie on a torus, as sometimes is the case in molecular biology and environmental sciences [31,9,10]. For data science applications, the results of [29] show that one can take data set and determine its homology with high degree of confidence. This would then inform one which M_0 would be appropriate for a diffeomorphic density matching problem for use in either redefining a mesh for computations or sampling from a desired distribution $\mu_1 \subset M$.

4.6. Example computations for the oblate sphere

The results from [11] state that even for smooth density functions f_0, f_1, the Optimal Transport mapping T with the squared geodesic cost function is not even guaranteed to be continuous. Therefore, using the cost function $\bar{c}$ on the unit sphere $\mathbb{S}^2$ is greatly preferable for performing diffeomorphic density matching computations using Optimal Transport.

We desire to solve the diffeomorphic mapping problem on the oblate sphere

$$\left\{(x,y,z) : x^2 + y^2 + 16z^2 = 1\right\}. \tag{55}$$

The diffeomorphism we choose is quite simple: $\Phi(x,y,z) = (x,y,4z)$ (note that this diffeomorphism is infinitely differentiable, but its inverse is only once differentiable). In local coordinates, we compute the Jacobian $d\Phi_z$. We see how $d\Phi_z$ acts on tangent elements, but choosing a path $\gamma : [0,1] \to M$ and computing:

$$d\Phi_z(\gamma'(0)) = (\Phi\circ\gamma)'(0). \tag{56}$$

We use a system of spherical polar coordinates on M. That is, $(x,y,z) = (\sin\theta\cos\phi, \sin\theta\sin\phi, 0.25\cos\theta)$ At a point $x \in M$, we use the local tangent coordinates given by the unit vectors $\hat{\theta}, \hat{\phi}$, where

$$\hat{\theta} = \frac{1}{\sqrt{\cos^2\theta + \frac{1}{16}\sin^2\theta}}\left(\cos\theta\cos\phi, \cos\theta\sin\phi, -\frac{1}{4}\sin\theta\right), \hat{\phi} = (-\sin\phi, \cos\phi, 0). \tag{57}$$

Start with a path just in the $\hat{\phi}$ direction $\gamma_1(t) = (\sin\theta_0\cos(t_0+t), \sin\theta_0\sin(t_0+t), 0.25\cos\theta_0)$. Then, $\gamma_1'(0) = (-\sin\theta_0\sin t_0, \sin\theta_0\cos t_0, 0)$, which has magnitude $\sin\theta_0$. After the diffeomorphism, we have $\Phi\circ\gamma_1(t) = (\sin\theta_0\cos(t_0+t), \sin\theta_0\sin(t_0+t), \cos\theta_0)$. Therefore,

$$(\Phi\circ\gamma_1)'(0) = (-\sin\theta_0\sin t_0, \sin\theta_0\cos t_0, 0), \tag{58}$$

which has magnitude $\sin\theta_0$. Let's define another path $\gamma_2(t) = (\sin(t_0+t)\cos\phi_0, \sin(t_0+t)\sin\phi_0, 0.25\cos(t_0+t))$ whose derivative is $\gamma_2'(t) = (\cos(t_0+t)\cos\phi_0, \cos(t_0+t)\sin\phi_0, -0.25\sin(t_0+t))$, which has magnitude $\sqrt{\cos^2 t_0 + \frac{1}{16}\sin^2 t_0}$. Then,

$$(\Phi\circ\gamma)'(0) = (\cos t_0\cos\phi_0, \cos t_0\sin\phi_0, -\sin t_0), \tag{59}$$

which has magnitude $\sqrt{\cos^2\theta_0 + \sin^2\theta_0} = 1$. Therefore, the linear map $d\Phi_z$ takes vectors in the $\hat{\phi}$-direction and stretches them by a factor of 1, and takes vectors in the $\hat{\theta}$-direction and stretches them by a factor of $1/\sqrt{\cos^2\theta_0 + \frac{1}{16}\sin^2\theta_0}$. Hence,

$$\left|d\Phi_z\right| = \frac{1}{\sqrt{\cos^2\theta_0 + \frac{1}{16}\sin^2\theta_0}}. \tag{60}$$

Therefore, in order to solve the Optimal Transport, we redefine the mass density functions $\bar{f}_0, \bar{f}_1 : M \to \mathbb{R}$ as:

$$f_i(\theta,\phi) = \bar{f}_i(\Phi^{-1}(\theta,\phi))\sqrt{\cos^2\theta + \frac{1}{16}\sin^2\theta} = \bar{f}(\theta,\phi)\sqrt{\cos^2\theta + \frac{1}{16}\sin^2\theta}. \tag{61}$$

5. Numerical methods for Optimal Transport and Optimal Information Transport

In this section, we propose using provably convergent methods for solving the Optimal Transport and Optimal Information Transport problem on the unit sphere, as developed in the line of work [15,16]. The first numerical scheme developed to solve the Optimal Transport problem on the sphere with the squared geodesic cost function was from [38] and continued in a line of work, including a finite-element method in [24]. The scheme in [38] relied on computing a fixed-point iteration from a linearization of an equidistribution equation on $\mathbb{S}^2$, which corresponds to the Optimal Transport problem with squared geodesic cost. This method was not furnished with a convergence proof. Nevertheless, the methods worked very well in various computational examples.

Recently, it came to the author's attention that numerical methods were also proposed for solving mean-field games on manifolds in [40], which included the Benamou-Brenier formulation of Optimal Transport as a subcase.

It should also be mentioned, as was stated in Section 2, that any choice of cost function that guarantees the regularity results in Theorem 3 would be appropriate for diffeomorphic density matching. This means that the entire line of numerical work that computes the reflector shape for the reflector antenna problem (using a logarithmic cost function) is appropriate (since the mapping T can be constructed from the reflector shape) to use for any genus-0 manifold, following the reformulation of the problem outlined in Section 4. The author is unaware of any numerical studies for the moving mesh problems using costs like the logarithmic cost arising in the reflector antenna problem, but wishes to bring to the attention of the computational community the viability of performing such computations.

The discretizations presented here are designed to be consistent and monotone, see [1,30] for more detail on the definition of monotonicity in finite-difference schemes. The properties of consistency, monotonicity, and Lipschitz stability are sufficient to guarantee convergence even in nonsmooth cases [15]. Furthermore, these are direct discretizations of the PDE (8), and do not rely on performing a linearization. They are also flexible enough to easily work with other cost functions, such as the logarithmic cost arising in the reflector antenna problem. Thus, they are well-suited to directly discretize the PDE (8) for any appropriate cost function (see [20]) and have convergence guarantees.

5.1. Grid, computational neighborhoods, & discrete derivatives

We describe the grid and discretization on the unit sphere, since computations on genus-0 surfaces can be reduced to computations on the unit sphere, as shown in Section 4. We start with a finite set of N grid points $\mathcal{G} \subset \mathbb{S}^2$, that have some given connectivity $\mathcal{K}$ (usually given as an adjacency matrix). This connectivity matrix $\mathcal{K}$ will remain unchanged after applying the computed mapping T to the grid $\mathcal{G}$. We let $d_{\mathbb{S}^2}(x, y)$ denote the usual geodesic distance between points x, y on the unit sphere $\mathbb{S}^2$. To the grid $\mathcal{G}$, we associate a number h which serves as a parameter indicating the overall spacing of grid points. More precisely,

$$h = \sup_{x \in \mathbb{S}^2} \min_{y \in \mathcal{G}^h} d_{\mathbb{S}^2}(x, y) = \mathcal{O}\left(N^{-1/2}\right). \tag{62}$$

This discretization parameter is the minimum radius h on the sphere that guarantees that the ball of radius h contains at least one discretization point. The grid must sufficiently resolve the sphere and satisfy some technical regularity requirements which are not difficult in practice to achieve, as detailed in [15]. These regularity requirements ensure that the grid is more or less isotropic with respect to direction.

5.1.1. Computation of second- and first-order directional derivatives

Both the Optimal Transport and Optimal Information Transport PDE involve a PDE computation involving the eigenvalues of a Hessian matrix of a function and the gradient of a function. In the former, the PDE (8) has a determinant of a Hessian matrix, which is equal to the product of the eigenvalues of the Hessian matrix. In the latter, the PDE (28) involves the computation of a Laplacian, which is the sum of the eigenvalues of a Hessian matrix. The eigenvalues of a Hessian matrix are second directional derivatives. Of course, the gradient is simply a first directional derivative in the tangent direction of steepest ascent.

To perform computations of the first and second directional derivatives at x_i, we begin by projecting grid points within a radius r of x_i onto the tangent plane at x_i. That is, we consider the set of relevant neighboring discretization points

$$\mathcal{Z}(x_i) = \left\{ z = \text{Proj}(x; x_i) \mid x \in \mathcal{G}, d_{\mathbb{S}^2}(x, x_i) \leq r \right\}. \tag{63}$$

The choice of the radius of the neighborhood is $r = \mathcal{O}(\sqrt{h})$ is optimal for the second directional derivative, see [12]. We use the same radius for our gradient computations.

The projection onto local tangent planes is accomplished using geodesic normal coordinates, which are chosen to preserve the distance from x_i (i.e. $d_{\mathbb{S}^2}(x, x_i) = \|x_i - \text{Proj}(x; x_i)\|$). This prevents any distortions that a non-zero curvature of a manifold introduces to second-order derivatives. This allows for monotone approximation schemes to be more easily constructed, which is our goal, since this will lead to a guarantee of convergence to the true solution. On the unit sphere, there are explicit formulas for geodesic normal coordinates about a point $x \in \mathbb{S}^2$. If, as mentioned in Section 4 we are performing computations for a genus-1 manifold, then, it is only necessary to have a consistent approximation for geodesic coordinates at a point x, since explicit formulas for geodesic normal coordinates do not exist in general. However, the case of genus-1 manifolds will not be treated in this manuscript, but will be deferred to future work.

When computing second and first-order directional derivatives, we will consider the following finite set of possible directions in the tangent plane (in geodesic normal coordinates),

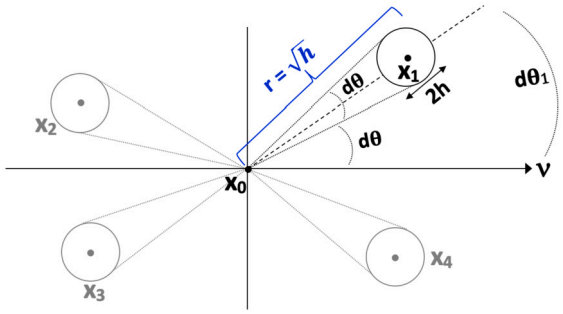


Fig. 1. The computational points are chosen so as to ensure that they are within $2d\theta$ of the axis of computation, but also stay at least $d\theta$ away from the axis to ensure monotonicity of the scheme.

$$V = \left\{ \{(\cos(jd\theta), \sin(jd\theta)), (-\sin(jd\theta), \cos(jd\theta))\} \mid j = 1, \dots, \frac{\pi}{2d\theta} \right\}, \tag{64}$$

where the angular resolution $d\theta = \dfrac{\pi}{2\lfloor \pi/(2\sqrt{h}) \rfloor}$.

In order to build approximations for second directional derivatives in each direction $v \in V$, we will select four grid points $x_j \in \mathcal{Z}(x_i)$, $j = 1, \dots, 4$, which will be used to construct the directional derivatives in this direction. To accomplish this, we let $v^\perp$ be a unit vector orthogonal to v and represent points in $x \in \mathcal{Z}(x_i)$ using (rotated) polar coordinates (r, θ) centered at x_i via

$$x = x_i + r(v\cos\theta + v^\perp \sin\theta), x \in \mathcal{Z}(x_i).$$

Then we select four points, each in a different quadrant $(Q_1, \dots, Q_4)$, that are well-aligned with the direction of v and approximately symmetric about the axis v via

$$x_j \in \underset{x \in \mathcal{Z}(x_i)}{\operatorname{argmin}} \left\{ |\sin\theta| \,|\, |\sin\theta| \geq d\theta, r \geq \sqrt{h} - 2h, x \in Q_j \right\}, \tag{65}$$

where $\cos\theta \geq 0$ for points in Q_1 or Q_4 and $\sin\theta \geq 0$ for points in Q_1 or Q_2, see Fig. 1.

From here, we construct approximations of second directional derivatives of the form

$$\mathcal{D}_{vv} u(x_i) = \sum_{j=1}^{4} a_j (u(x_j) - u(x_i)). \tag{66}$$

Denote θ_i to be the angle that x_i makes with the axis v. Also, denote h_i, as the distance from x_i to x_0. The work in [12] shows that the optimal choice is $h_i, \theta_i = \mathcal{O}(\sqrt{h})$. After performing a Taylor expansion of the sums in (66) we get explicit values for a_j. The Taylor expansion proceeds as follows:

$$\sum_{j=1}^{4} a_j (u(x_j) - u(x_i)) = \sum_{i=1}^{4} a_i \left(h_i \cos\theta_i u_v(x_0) + h_i \sin\theta_i u_v(x_0) + \dots \right). \tag{67}$$

In order to approximate the second directional derivative in the v direction, this requires that $\sum_{i=1}^{4} \frac{1}{2} a_i h_i^2 \cos^2\theta_i = 1$. This indicates that $a_i = \mathcal{O}(h^{-1})$. Thus, we find the asymptotic ordering:

$$\begin{aligned} a_i h_i \cos\theta_i &= \mathcal{O}(h^{-1/2}), \\ a_i h_i \sin\theta_i &= \mathcal{O}(1), \\ a_i h_i^2 \cos^2\theta_i &= \mathcal{O}(1). \end{aligned} \tag{68}$$

The system is augmented by another condition $a_1 h_1 \sin\theta_i + a_4 h_4 \sin\theta_4 = 0$ and the coefficients a_i are given explicitly in [12]. They, importantly, satisfy $a_i \geq 0$, allowing us to build monotone discretizations.

One possible discretization for the gradient, again guided by the work in [12], suggests that we choose $r = \mathcal{O}(\sqrt{h})$ and consider 4 points, where one point is in each quadrant $(Q_1, \dots, Q_4)$, and are well-aligned with the direction of computation v and approximately symmetric about the axis v, as above for the second-directional derivatives. We approximate first directional derivatives in the direction v as:

$$\mathcal{D}_v u(x_0) \approx \sum_{i=1}^{8} b_i \left(u(x_i) - u(x_0) \right). \tag{69}$$

Again, denote θ_i to be the angle that x_i makes with the axis v, which is $\mathcal{O}(\sqrt{h})$, by [12]. Also, denote h_i, which is also $\mathcal{O}(\sqrt{h})$ as the distance from x_i to x_0. Then,

$$\sum_{i=1}^{8} b_i \left(u(x_i) - u(x_0)\right) = h_i \cos\theta_i u_{\nu}(x_0) + h_i \sin\theta_i u_{\nu_\perp}(x_0) + \dots \tag{70}$$

By the equation $\sum_{i=1}^{8} b_i h_i \cos\theta_i = 1$, we see that $b_i = \mathcal{O}(h^{-1/2})$. Thus, we get the asymptotic ordering:

$$\begin{aligned} b_i h_i \cos\theta_i &= \mathcal{O}(1), \\ b_i h_i \sin\theta_i &= \mathcal{O}(\sqrt{h}), \\ b_i h_i^2 \cos^2\theta_i &= \mathcal{O}(\sqrt{h}). \end{aligned} \tag{71}$$

Thus, an approximation with consistency error $\mathcal{O}(h)$ is furnished by solving the following system of linear equations for b_i:

$$\begin{aligned} \sum_{i=1}^{8} b_i h_i \cos\theta_i &= 1, \\ \sum_{i=1}^{8} b_i h_i \sin\theta_i &= 0, \\ \sum_{i=1}^{8} b_i h_i^2 \cos^2\theta_i &= 0, \\ b_1 h_1 \sin\theta_i + b_4 h_4 \sin\theta_4 &= 0 \end{aligned} \tag{72}$$

And the approximation for the gradient at x_i is thus,

$$\nabla^h u^h(x_i) = \max_{\nu \in V} \mathcal{D}_{\nu} u(x_i) \tag{73}$$

5.2. Discretization of Optimal Transport and Optimal Information Transport PDE

The full monotone discretization for the discrete operator F^h for the Optimal Transport PDE (8) on the sphere is (see [16] for more details and analysis):

$$F^h\left(x_i, u^h(x_i)\right) = \min_{\{\nu_1,\nu_2\}\in V} \prod_{j=1}^{2} \max\left\{ \mathcal{D}^h_{\nu_j\nu_j} u^h(x_i) + g^-_{1,\nu_j}\left(\nabla^h u^h(x_i)\right), 0 \right\} - f_1(x_0) g_2^+\left(\nabla^h u^h(x_i)\right), \tag{74}$$

where $g_1(p;\nu) = -\mathcal{D}^h_{\nu\nu} c(x_i, y)|_{y=T(x_i,p)}$ and $g_2(p;\nu) = \frac{\sin\|p\|}{\|p\|}$ and $g^{\pm}_{i,\nu}\left(\nabla^h u^h(x_i)\right) = g_i\left(\nabla^h u^h(x_i);\nu\right) \mp \epsilon_g \Delta^h u^h(x_i)$.

And the full monotone discretization for the Laplacian operator F^h solving the Poisson equation in (28) on the sphere is:

$$F^h\left(x_i, u^h(x_i)\right) = \Delta^h u^h(x_i) + \epsilon^h u^h(x_i) = f(x_i), \tag{75}$$

where $\Delta^h u^h(x_i) = \mathcal{D}^h_{\nu} u^h(x_i) + \mathcal{D}^h_{\nu_\perp} u^h(x_i)$, for any $\nu \in V$ and $\epsilon^h \to 0$ is chosen to be the consistency error of the operator Δ^h.

5.3. Convergence results

The full details of the convergence proof and scheme for solving the Optimal Transport PDE (8) with a consistent and monotone discretization with an additional discrete Lipschitz control are given in [15,16]. The result is that the computed solution u^h converges uniformly to the solution u of (8), but thus far no explicit rates of convergence have been established.

For the discretization of the Poisson equation (75), results on monotone and consistent discretizations of linear elliptic PDE in divergence form recently obtained in [17] show that there exist schemes for which the explicit convergence rate is at worst $\left|u^h - u\right| = \mathcal{O}\left(h^{\alpha/3}\right)$, where u are *a priori* C^3 solutions of the Poisson equation on M and $\mathcal{O}(h^{\alpha})$ is the discretization error of the Laplacian.

5.4. Algorithm for Optimal Transport

The procedure for solving the discrete Optimal Transport PDE (74) is by using a parabolic scheme presented in Algorithm 1.

5.5. Algorithm for Optimal Information Transport

In order to solve the Optimal Information Transport problem on M, we must solve the system of equations (28) for the diffeomorphic mapping T. First, we precompute θ from (20). Denote the forward map at a time step n will be denoted by T_n and the corresponding inverse map at a time step n will be denoted by S_n. Then, we update T_n and S_n via a forward Euler scheme with time step size Δt and iterate until $n\Delta t = 1$.

Algorithm 1 Computing the solution to elliptic PDE $F[u] = 0$.

1: Initialize u_0^h;
2: Fix $\epsilon > 0, \Delta t > 0$;
3: **while** $\left| F^h\left(x_i, u_n^h(x_i)\right)\right| > \epsilon$, where F^h is given by (74) **do**
4: Compute $u_{n+1}^h(x_i) = u_n^h(x_i) + \Delta t F^h\left(x_i, u_n^h(x_i)\right)$;
5: **end while**
6: Compute the gradient of u^h via $\nabla^h u^h(x_i)$ using (73).
7: Use the explicit expression (7) for the exponential map to compute the approximation to the mapping T^h.

The algorithm requires two important functions: Proj, a projection map, and Interp, a consistent interpolation map. For the unit sphere, we use explicit formulas for the exponential map for the projection map. Given a tangent vector $p \in T_x\mathbb{S}^2$, the explicit expression for the exponential map on the unit sphere is:

$$\exp_x(p) = x\cos(\|p\|) + \frac{p}{\|p\|}\sin(\|p\|) \tag{76}$$

The interpolation map Interp is of especial importance and the Matlab package 'griddata' was used.

Algorithm 2 Computing the diffeomorphic mapping $\varphi = S(1)$.

1: Initialize $T_0 = \mathrm{Id}$;
2: Initialize $S_0 = \mathrm{Id}$;
3: Fix $\Delta t \ll 1$;
4: Precompute θ via quadrature;
5: **while** $n\Delta t < 1$ **do**
6: Compute the density function $\nu_n := \ddot{\mu}_n/\mu_n$ using the explicit formulas on the grid $\left\{S(x_j)\right\}_j$;
7: Solve $\Delta^h f_n(x_i) = \nu_n\left(S_n(x_i)\right)$ for $f_n(x_i)$ with the monotone, consistent discretization of the Laplacian given in (75) and solve the resulting linear system of equations using standard linear algebra techniques;
8: Compute $\nabla^h f_n(x_i)$ for all x_i using (73);
9: Interpolate: $\nabla^h f_n(x_i)$ onto the grid $\left\{T(x_j)\right\}_j$;
10: Compute $T_{n+1}(x_i) = \mathrm{Proj}\left\{T_n(x_i) + \Delta t \nabla^h f_n\left(T_n(x_i)\right)\right\}$, using (76);
11: Compute $S_{n+1}(x_i) = S_n\left(\mathrm{Proj}\left\{x - \Delta t \nabla^h f_n(x_i)\right\}\right)$;
12: **end while**

For nonsmooth densities, we used a nearest-neighbor interpolation (piecewise discontinuous), in order to retain the discontinuities of the densities for more challenging computational examples, but otherwise used linear interpolation, especially for the mapping S and the gradient ∇f_n. It is of vital importance *not* to use a piecewise constant interpolation on ∇f_n and S, since otherwise for small enough Δt, the updates would remain unchanged.

6. Computational results

In this section, we show the results of diffeomorphic density matching computations using Optimal Transport and Optimal Information Transport. We provide evidence that in the absence of higher-order schemes for Optimal Transport, the preferred method should be Optimal Information Transport, due to the ease of implementation, speed, and accuracy of solutions. With the same finite-difference discretizations on unstructured grids of first- and second-order derivatives in both the Optimal Transport and Optimal Information Transport computations, we see improved regularity of the mapping with Optimal Information Transport.

As shown in Section 4, for manifolds diffeomorphic to the sphere, diffeomorphic density matching computations can always be performed on the sphere. Therefore, we are able to carry out an Optimal Information Transport computation of densities on the oblate sphere by modifying the problem appropriately to the unit sphere.

Our objective is to perform adaptive mesh computations on the sphere. Here we will be using a mesh defined on a cube which is then projected onto the sphere, see Fig. 2 for an example of the generation of such a mesh.

The original mesh, therefore, *does not have constant density of nodes*. However, in the moving mesh problem, we will be performing computations that produce diffeomorphic maps from a constant source density to a variable target density. This allows us to redistribute the mesh as desired treating the original cube mesh projected on the sphere as if it had constant density of nodes.

While a convergence theorem, such as in Section 5.3, gives convergence guarantees for the discretization of the Optimal Information Transport problem, it is difficult to furnish explicit solutions on manifolds and also difficult to define a residual. For both the Optimal Transport and Optimal Information Transport problems, the residual can be defined as the discretization of the following PDE operator:

$$H(x, T(x)) = f_0(x) - |DT(x)|\, f_1(T(x)), \tag{77}$$

where $|DT(x)|$ is the determinant of the Jacobian for the mapping. For the Optimal Transport problem, this has the PDE formulation (8) and thus the L^∞ residual can be computed simply using the discretization of the PDE (77) and then computing the maximum of the absolute value of this quantity over all computational points.

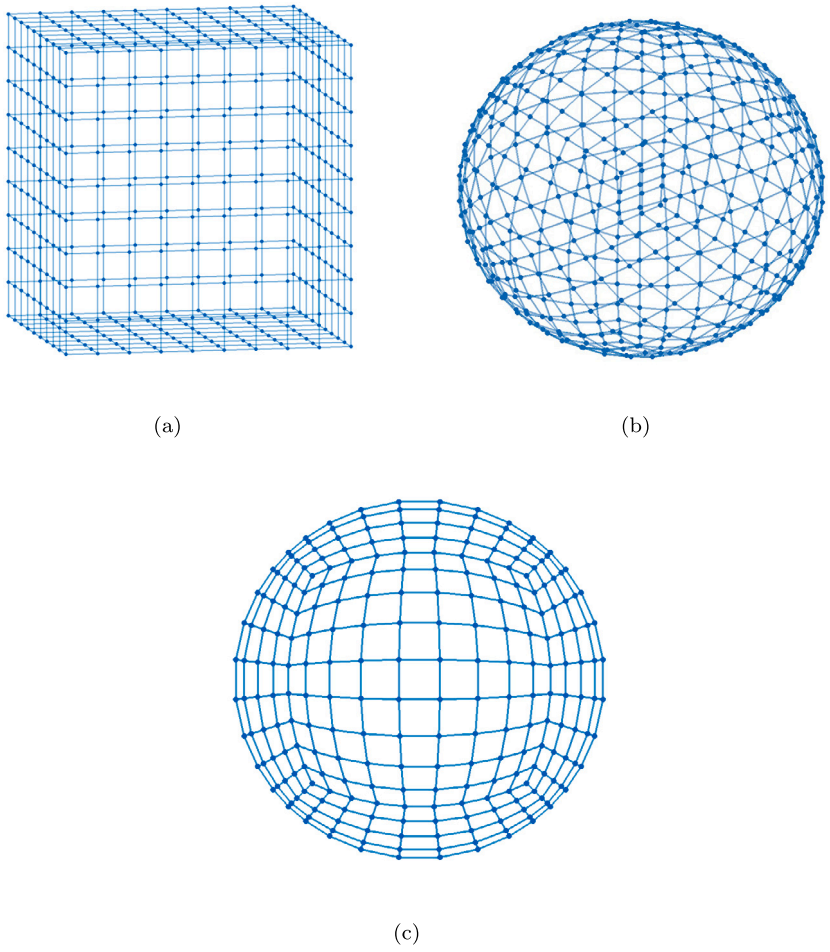


Fig. 2. A mesh is generated on a cube 2(a) which is then projected to the unit sphere to create a mesh on the sphere 2(b). The view of one hemisphere of the mesh on the sphere is shown in 2(c).

For the Optimal Information Transport problem, once we have run a computation, we can approximate the quantity $|DT|$ by computing the amplification factor of the area of a Voronoi cell about a given node after being transformed by the mapping. A workable definition for the residual is the average of the absolute value of the discretization of the quantity:

$$\tilde{H}(x, T(x)) = \frac{f_0(x)}{f_1(T(x))} - |DT(x)|, \tag{78}$$

provided that $f_1 \neq 0$. In practice, computing this residual is very inaccurate. In Section 6.3, we perform the computation for the Optimal Information Transport problem.

Exact solutions are possible to construct for the Optimal Transport problem on the sphere for many cost functions. We perform a computation for a solution known exactly in Section 6.3 for the squared geodesic cost function on the sphere. Constructing exact solutions for the Optimal Information Transport problem, on the other hand, is much harder. These computations would rely on one being able to construct paths $\varphi(t)$ which are already known to be horizontal geodesics. Thus, most work using Optimal Information Transport lacks traditional convergence plots. An investigation into using better residuals and computing exact solutions for Optimal Information Transport will be deferred to future work.

All computations were performed using Matlab R2021b on a 13-inch 2017 MacBook Pro, with a 2.3 GHz Dual-Core Intel Core i5 and 16 GB of 2133 MHz LPDDR3 memory. The grid sizes in the figures had $N = 5048$, $N = 10088$, or $N = 20888$ nodes. The Optimal Transport computations ran for about $10 - 60$ minutes for computations of size $N = \mathcal{O}(10^4)$ and the Optimal Information Transport computations took about 5 minutes for computations of comparable size. The computations could have been scaled up, but are limited in this manuscript since it was desirable to produce a mesh that could be seen in the figures and as proof of concept.

6.1. Equatorial node density concentration

First, we demonstrate that the computation of mesh redistribution using Optimal Transport and Optimal Information Transport can be both successful in producing meshes which do not exhibit tangling. We select density functions with the goal of producing a

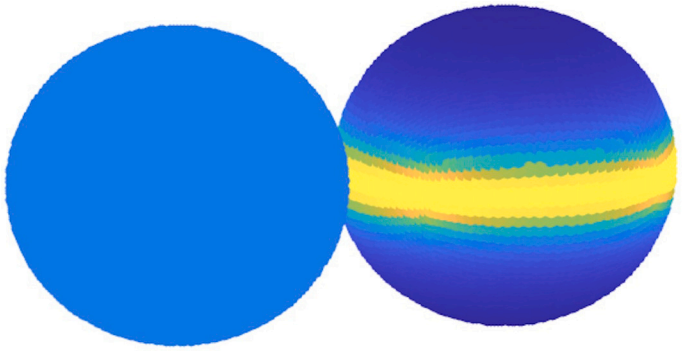


Fig. 3. Density in the original mesh (left) and the modified density (right).

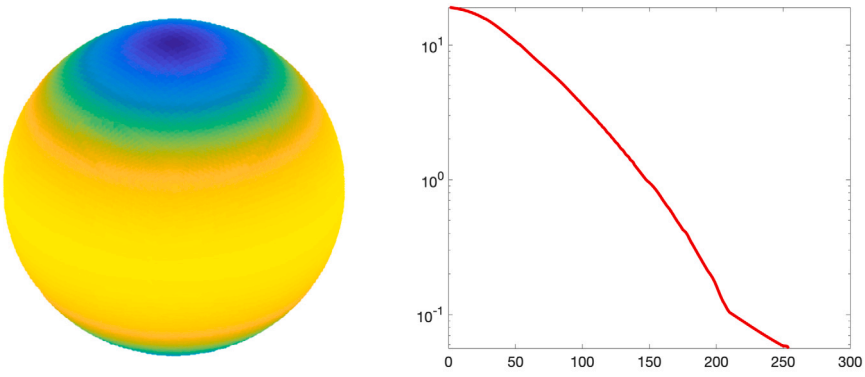


Fig. 4. The potential function u in Fig. 4(a) and the residual of the computation by iteration in Fig. 4(b).

transport map T that will concentrate mesh points around the equator, see (79). These f_0, f_1 are defined for the Optimal Transport problem. For the Optimal Information Transport problem, each will be multiplied by a factor of 4π so that $\int_{\mathbb{S}^2} f_0 dx = \int_{\mathbb{S}^2} f_1 dx = \text{vol}(\mathbb{S}^2)$.

$$\begin{aligned} f_0(\theta,\phi) &= \frac{1}{4\pi} \\ f_1(\theta,\phi) &= \left(1 - \exp\left(\frac{-1}{30\,(\theta-\pi/2)^2}\right)\right)/3.53552 \end{aligned} \tag{79}$$

We then require that the density concentrates about the equator according to (79), which is shown in Fig. 3. The Optimal Transport computation first yields the potential function u and the residual, see Fig. 3. Then, the mapping T is computed via the potential function, see Fig. 4, and the resulting mesh restructuring computed using Optimal Transport is pictured in Fig. 5.

Now we perform the same computation using Optimal Information Transport, see Fig. 6.

6.2. Oceanic node density concentration

Another example serves to demonstrate the versatility and desirability of using Optimal Information Transport over Optimal Transport for moving mesh problems. We take a constant source density $f_0 = 1$ and using Optimal Information Transport map it to a complicated, discontinuous map of the world, where a higher density of grid points is required over the oceans and lower density over the continents, see Fig. 7. From the end of Section 2.1, using Optimal Transport methods to find a mapping T to a discontinuous target density function only guarantees that the mapping T is at least continuous. There are no such guarantees for Optimal Information Transport, since the source and target densities must be smooth. Nevertheless, we proceed and try and apply these PDE and solve for the numerical solution anyway.

The result is shown in Fig. 8.

Some difficulties are encountered when attempting to perform the same numerical computation via Optimal Transport, where we instead choose f_0 to encode the globe density and $f_1 = 1/(4\pi)$ to be the constant density, see Fig. 9. Having solved for the potential function, see Fig. 10, we still have significant tangling for very complicated images like the world map, see Fig. 11.

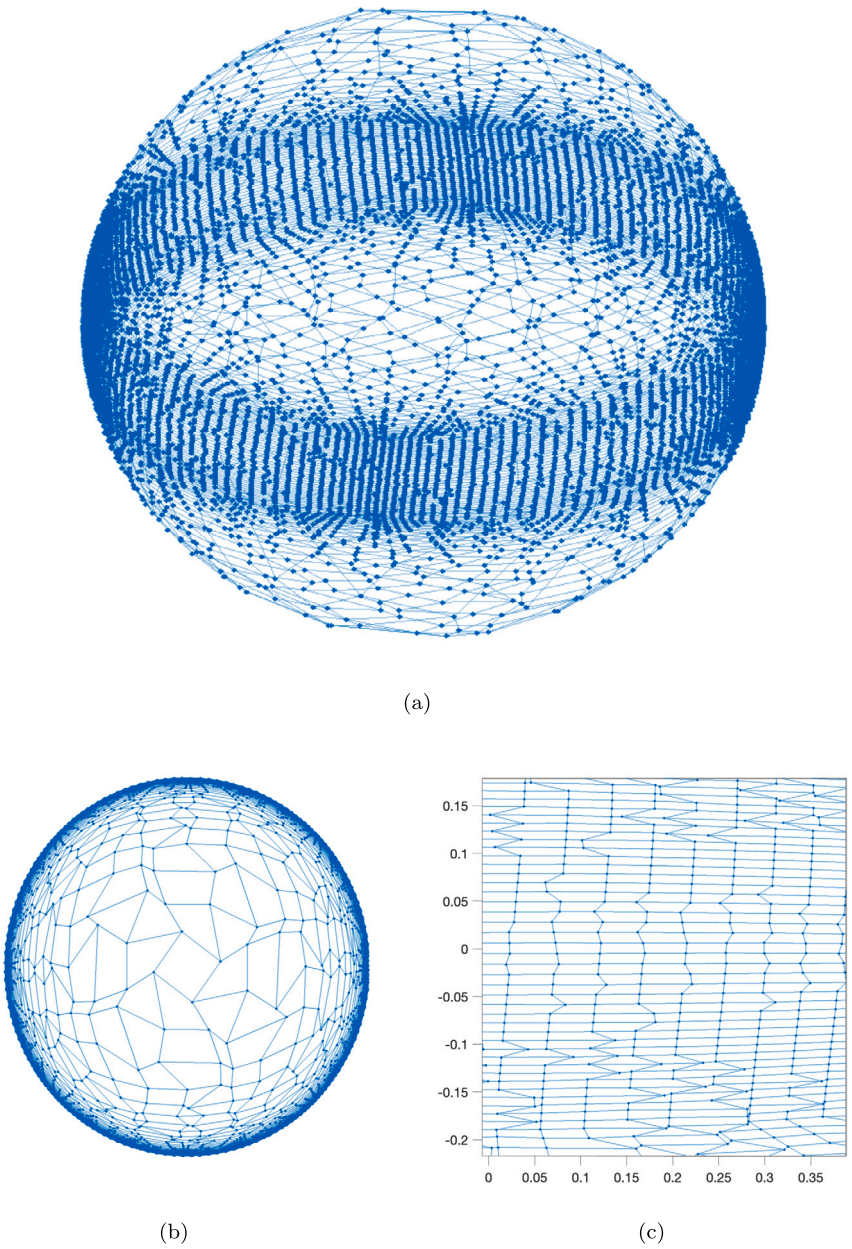


Fig. 5. In Fig. (a) we have 3D view of the mesh obtained via Optimal Transport map, in Fig. (b) a top view, and in Fig. (c) a detailed view of the mesh around the equator showing that the grid lines do not tangle, with $N = 10088$ nodes.

One potential solution is to develop higher-order discretizations of the Optimal Transport PDE and will be explored in future work.

6.3. Computations for explicit solutions

Following the derivation in [6], if the Optimal Transport potential function is axi-symmetric, that is, if $u(\theta, \phi) = u(\theta)$, then the mapping $T(\theta, \phi) = (\theta', \phi) = \left(\theta + \frac{du}{d\theta}, \phi\right)$. Using this, we can compute the determinant of the Jacobian of the Optimal Transport mapping T.

$$F(\theta) := |DT| = \frac{\sin\theta'}{\sin\theta}\frac{d\theta'}{d\theta}. \tag{80}$$

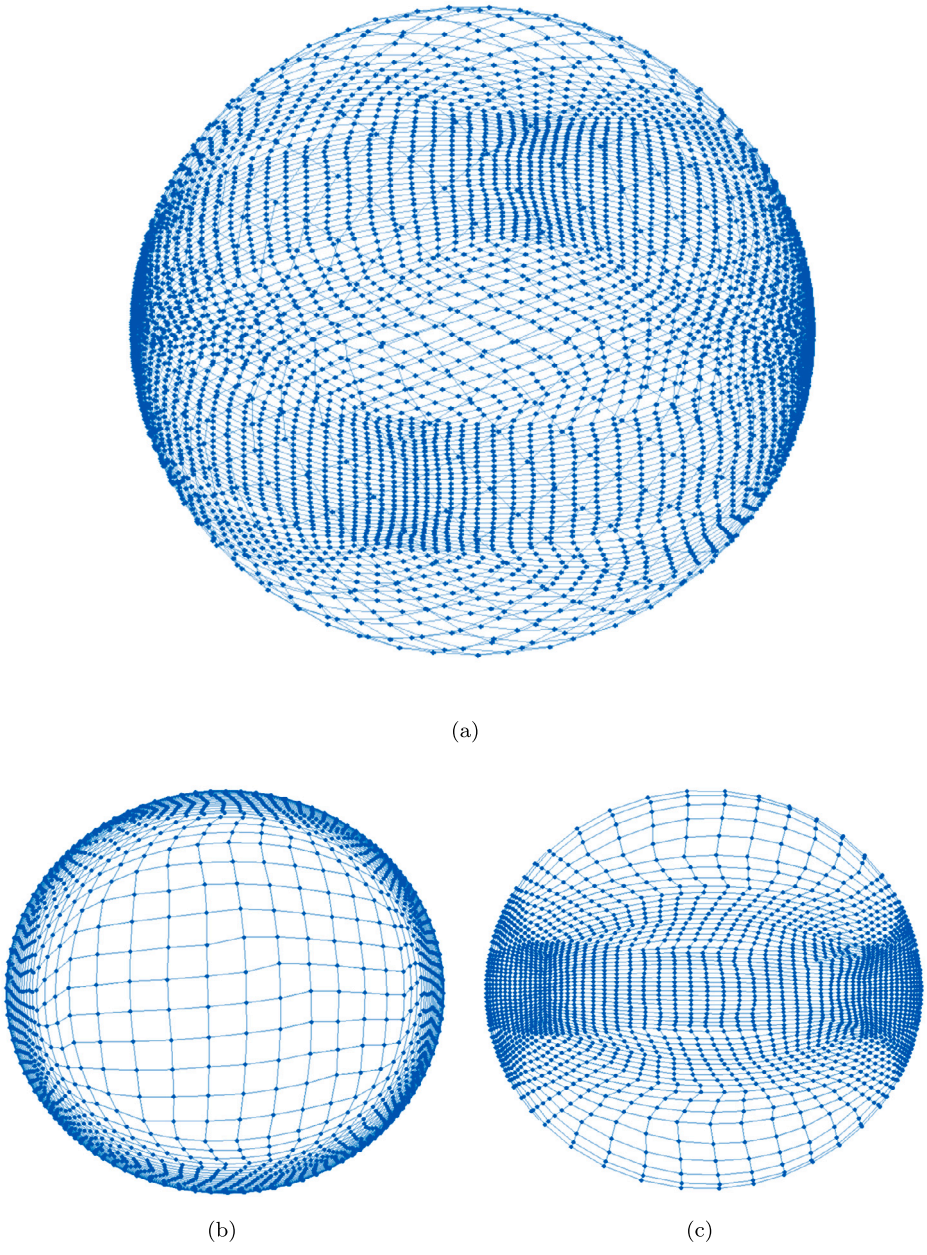


Fig. 6. Fig. (a) shows the 3D view of the mesh obtained via Optimal Information Transport, Fig. (b) shows the top view, and a detailed view of the equator is shown in Fig. (c), demonstrating that the grid lines do not tangle and the solution has nice regularity properties with $N = 10088$ nodes.

For the squared geodesic cost function, this becomes

$$F(\theta) = \frac{\sin\left(\theta + \frac{du}{d\theta}\right)}{\sin\theta} \frac{d}{d\theta}\left(\theta + \frac{du}{d\theta}\right) = \frac{\sin\left(\theta + \frac{du}{d\theta}\right)}{\sin\theta}\left(1 + \frac{d^2u}{d\theta^2}\right). \tag{81}$$

We restate here that diffeomorphic density matching can also be achieved using the logarithmic cost function $c(x, y) = -\log(1 - x \cdot y)$ arising in the reflector antenna problem, see [36,37]. Therefore, it is possible to compute the potential function u using numerical methods for the reflector antenna problem. If so, the factor F will be different, because $\theta' = \theta - \frac{2}{\left(\frac{du}{d\theta}\right)^2 + 1}$. In this case, we derive a slightly more complicated expression:

$$F(\theta) = \frac{\sin\left(\theta + \arccos R\left(\frac{du}{d\theta}\right)\right)}{\sin\theta}\left(1 + \frac{d}{d\theta}\arccos R\left(\frac{du}{d\theta}\right)\right), \tag{82}$$

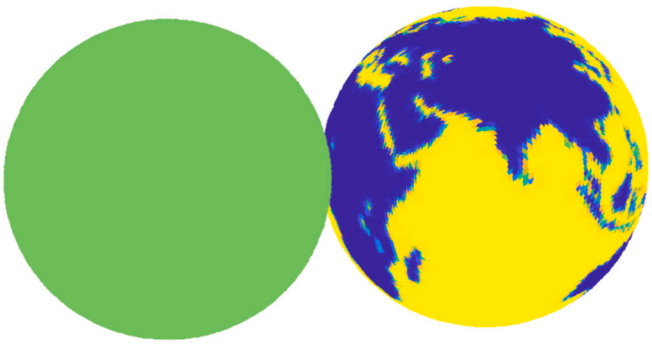


Fig. 7. Constant density in the source density (left) and the target globe density (right) concentrating mesh nodes in the oceans.

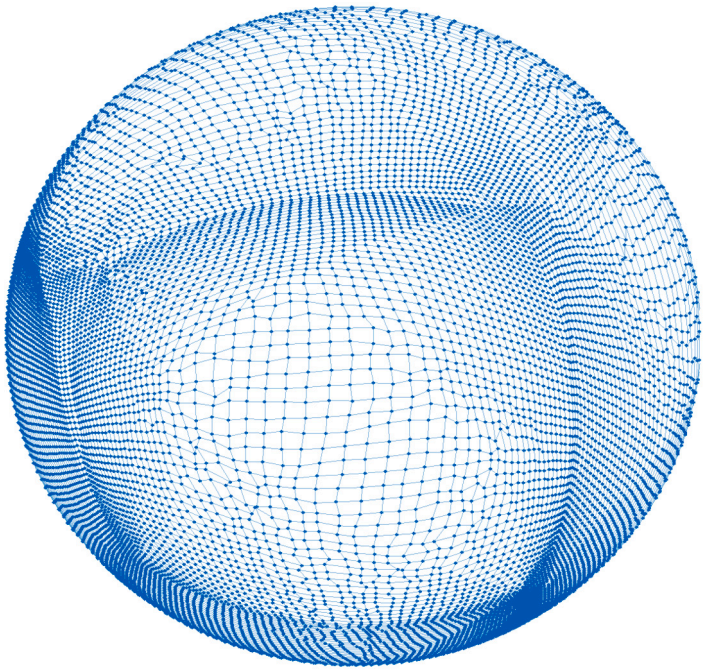


Fig. 8. Spherical mesh after redistribution of the density of nodes via Optimal Information Transport. The view is of North and South America, where the density of the mesh is less than the density of the mesh over the oceans with $N = 20888$ nodes.

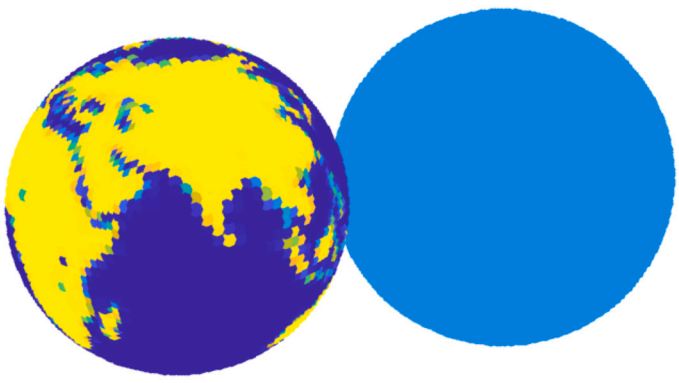


Fig. 9. Source globe density (right) concentrating mesh nodes in the continents and constant target density (right).

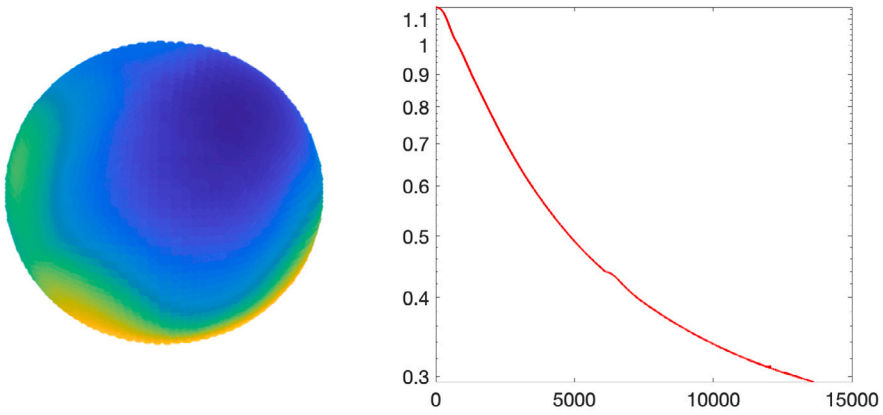


Fig. 10. The potential function u in Fig. 10(a) and the residual of the computation by iteration in Fig. 10(b).

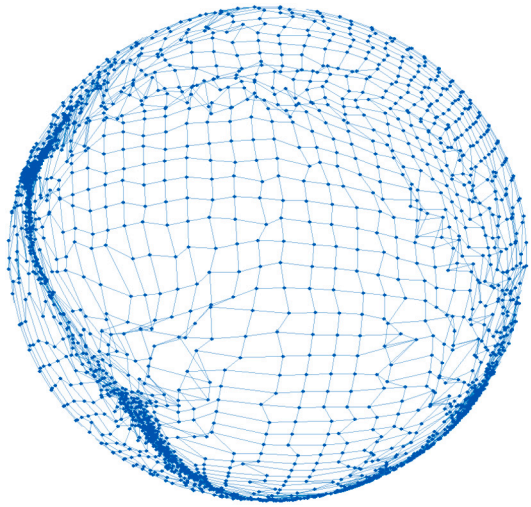


Fig. 11. The mesh after redistribution of the density of nodes using Optimal Transport. Detail on Africa, the Middle East, and Asia with other side of sphere hidden from view in order to show the mesh more clearly. One can observe some tangling with $N = 5048$ nodes.

where

$$R(\xi) = \frac{\xi^2 - 1}{\xi^2 + 1}. \tag{83}$$

Returning to the squared geodesic cost function, in contrast to the paper [6], we are concerned not with the monitor function, but with producing the density functions f_0 and f_1 given the determinant of the Jacobian of the mapping in (81). In order to have such an axi-symmetric mapping, it can easily be shown that the density functions f_0 and f_1 must also be axi-symmetric, i.e. $f_0(\theta, \phi) = f_0(\theta)$ and $f_1(\theta, \phi) = f_1(\theta)$. Furthermore, the measure-preserving property implies that f_0 and f_1 must satisfy the following Jacobian equation:

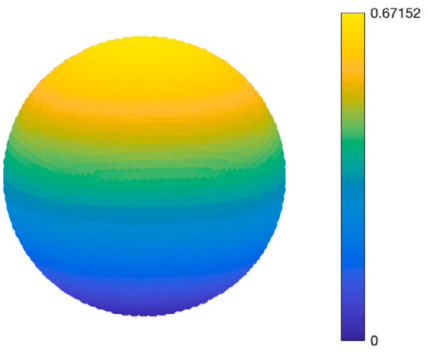
$$f_0(\theta) = F(\theta) f_1\left(\theta + \frac{du}{d\theta}\right). \tag{84}$$

For simplicity and ease of computation, let us take the target density to be constant $f_1 = 1/4\pi$ and we will solve for the source density function f_0 given a known smooth potential function u. The density function f_0 must satisfy

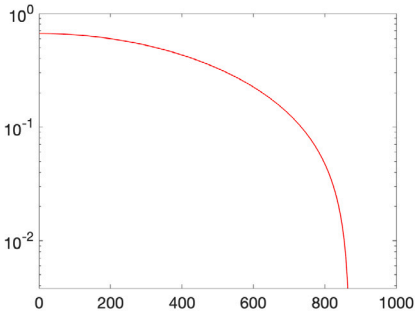
$$f_0(\theta) = \frac{F(\theta)}{4\pi}. \tag{85}$$

In this subsection, we will perform computations for the known explicit solution $u = \frac{\cos\theta}{a_0}$, where $a_0 \in \mathbb{R}$, $a_0 \neq 0$. Given this particular potential function, we have

(a)



(b)



(c)

Fig. 12. The densities f_0 and f_1 in Fig. 12(a) yielding the potential function $u = z/a_0$ in Fig. 12(b). The convergence of the error is shown in 12(c), as a function of iterations.

$$f_0(\theta) = \frac{\sin\left(\theta - \frac{\sin\theta}{a}\right)}{4\pi \sin\theta}\left(1 - \frac{\cos\theta}{a}\right). \tag{86}$$

This integrates to 1 over $\mathbb{S}^2$, by Equation (84). We can perform an Optimal Transport computation using Algorithm 1 with the choice $a_0 = 3$ yielding the densities and potential function in Fig. 12, with an L^∞ error equal to 0.0038 using $N = 5048$ nodes. Note, that the potential function we plot here is $v := u - \min u = \frac{1}{a_0}(z+1)$. The resulting mesh redistribution is shown in Fig. 13.

In order to perform the same computation using Optimal Information Transport, we take $f_0 = \frac{1}{4\pi}$ and $f_1 = C_0 \frac{\sin\theta}{\sin\left(\theta - \frac{\sin\theta}{a}\right)\left(1 - \frac{\cos\theta}{a}\right)}$, where $C_0 > 0$ is chosen such that $\int_{\mathbb{S}^2} f_1(x)dx = 1$. There is no analogue to the Optimal Transport potential function for Optimal Information Transport, but we can compute the residual by taking the maximum of Equation (78) over all nodes. We restate here that a search for better residuals is deferred to future work. The result is a residual error of 0.1929 computed via (78) and the computation is shown in Fig. 14.

6.4. Optimal Information Transport computation with significant compression

Here we show an example where there is significant compression. The target density we use is:

$$f_1(\theta,\phi) = (1-\epsilon)\frac{4\pi}{1.57949} e^{-30(\arccos(\sin\theta\cos\phi)-0.9)^2} + \epsilon \tag{87}$$

where $\epsilon = 0.05$, see Fig. 15. Using Optimal Information Transport, we compress the mesh severely around a ring, the result we can see in Fig. 16.

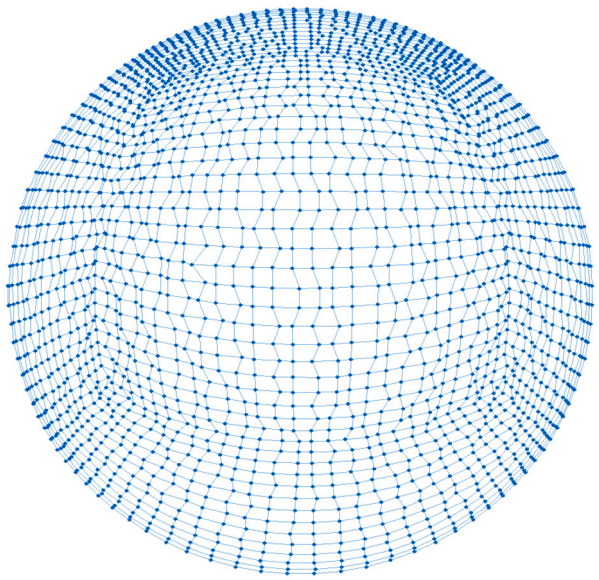


Fig. 13. The mesh redistribution via densities f_0 and f_1 via an Optimal Transport computation. As expected, we see a decrease in concentration at the south pole and an increase in concentration at the north pole, since the ratio f_1/f_0 is less than 1 at the south pole and greater than 1 at the north pole.

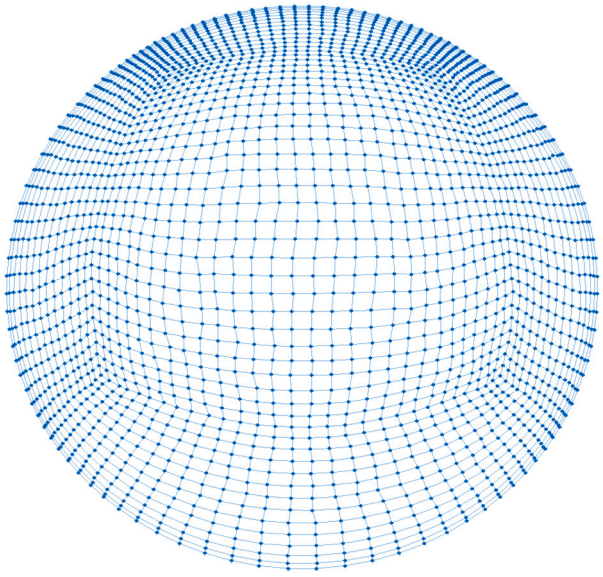


Fig. 14. The mesh redistribution via Optimal Information Transport. As expected, we see a decrease in concentration at the south pole and an increase in concentration at the north pole and the grid is more regular.

6.5. Optimal Information Transport computation on an oblate sphere

Here, we perform a computation on the oblate sphere $M = \left\{(x, y, z) : x^2 + y^2 + 16z^2 = 1\right\}$ using Optimal Information Transport to show the success of the method, following the derivation in Section 4. We desire to change the density of nodes on the oblate sphere by the following ratio of densities, as shown in Fig. 17:

$$\bar{f} = \text{vol}(M)\left(1 - e^{\frac{-1}{10\theta^2}}\right)/0.878646. \tag{88}$$

As shown in 4, we can modify the density of nodes in the same way by instead finding the pushforward T between the following density functions on $\mathbb{S}^2$:

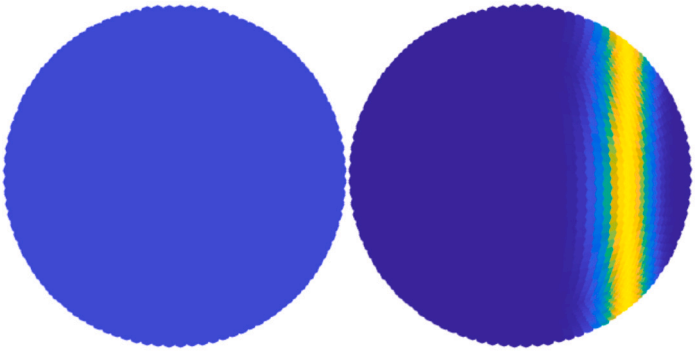


Fig. 15. A constant density mapping to a significantly compressed density.

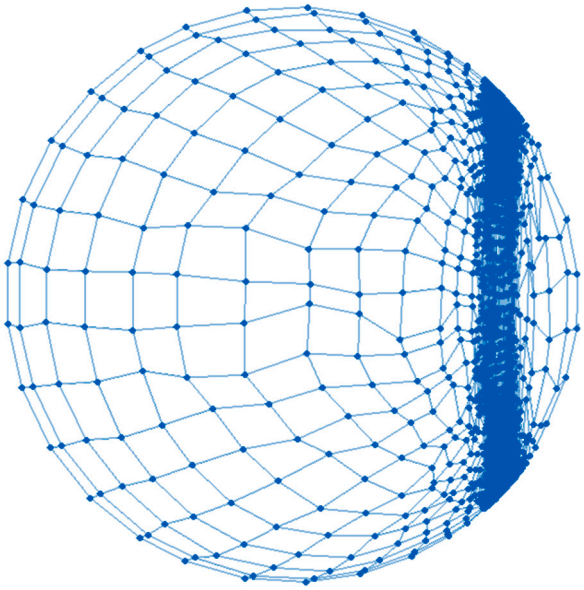


Fig. 16. A significantly compressed mesh with $N = 5048$ nodes.

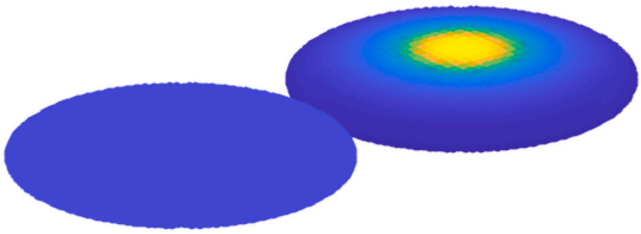


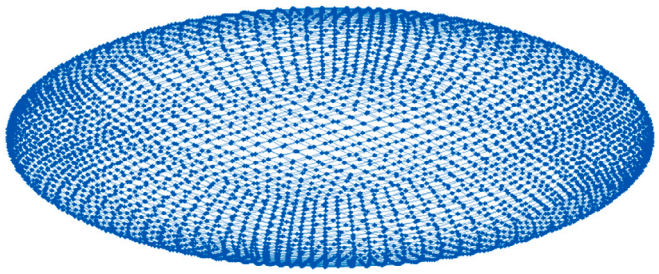
Fig. 17. Constant density in the source density (left) and the target density (right) concentrated about the north pole of the oblate sphere.

$$\begin{aligned} f_0 &= 1, \\ f_1 &= 4\pi \left(1 - e^{\frac{-1}{10\theta^2}}\right) / 0.878646. \end{aligned} \tag{89}$$

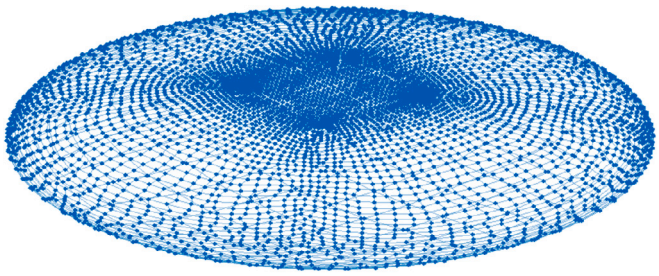
We perform the computation on $\mathbb{S}^2$ to find the new mesh, as shown in Fig. 18 and Fig. 19.

7. Conclusion

We have provided a comparison of Optimal Transport and Optimal Information Transport on the sphere for the diffeomorphic density matching problem with applications to adaptive mesh methods by using convergent finite-difference schemes and in doing so identified a much broader set of Optimal Transport cost functions which would be appropriate for moving mesh methods. In the

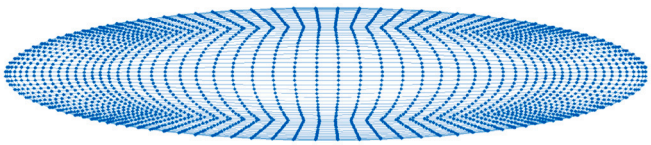


(a)

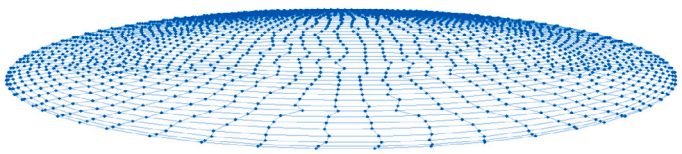


(b)

Fig. 18. Redistribution of mesh on an oblate sphere using Optimal Information Transport. The mesh before redistribution of nodes is in Fig. 18(a) and the result of the restructured mesh is in Fig. 18(b) with $N = 5048$ nodes.



(a)



(b)

Fig. 19. A side view shows the original mesh in 19(a) and the effect of the increase of the density of nodes on one side of the oblate sphere and decrease of the density of nodes on the other side in 19(b).

Optimal Information Transport problem, we solve a Poisson equation, compute a gradient, perform projections (via the exponential map), and perform interpolations. The choice of interpolation has a lot of bearing on the success of the computation. However, the output of the Optimal Information Transport computation is a mapping, whereas the output of the Optimal Transport problem is a potential function which then needs to undergo a further computation in order to yield a mapping. In order for Optimal Transport computations for diffeomorphic density matching to compete with Optimal Information Transport computations, it will be necessary to develop higher-order schemes, which will be pursued in further research.

We also showed how to reduce the diffeomorphic density matching problem on manifolds M that are diffeomorphic (i.e. there exists a diffeomorphism $\Phi : M \to \mathbb{S}^2$) to the sphere by solving a related problem on the sphere. Consequently, by denoting the squared geodesic cost between $x, y, \in \mathbb{S}^2$ by $c(x, y)$, we derived regularity guarantees for the Optimal Transport problem on M with the cost function $c(\Phi(z), \Phi(\xi))$, for $z, \xi \in M$. Future work will explore diffeomorphic density matching computations on genus-1 compact and connected surfaces with applications, by using a fixed "standard" torus upon which all computations will be performed.

CRediT authorship contribution statement

Axel G.R. Turnquist: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Axel G R Turnquist reports financial support was provided by National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

I would like to express my gratitude to Brittany Froese Hamfeldt for fruitful discussions, advice, and encouraging me to write this manuscript. The author was partially supported by an NSF GRFP 1849508 and by NSF DMS 1751996.

References

[1] G. Barles, P.E. Souganidis, Convergence of approximation schemes for fully nonlinear second order equations, Asymptot. Anal. 4 (1991) 271–283.
[2] Martin Bauer, Sarang Joshi, Klas Modin, Diffeomorphic density matching by optimal information transport, SIAM J. Imaging Sci. 8 (3) (2015) 1718–1751.
[3] Martin Bauer, Sarang Joshi, Klas Modin, Diffeomorphic random sampling using optimal information transport, in: Geometric Science of Information 2017, 2017, pp. 135–142.
[4] C.J. Budd, M.J.P. Cullen, E.J. Walsh, Monge-Ampère based moving mesh methods for numerical weather prediction, with applications to the Eady problem, J. Comput. Phys. 236 (2013) 247–270.
[5] C.J. Budd, J.F. Williams, Moving mesh generation using the parabolic Monge-Ampère equation, SIAM J. Sci. Comput. 31 (5) (2009) 3438–3465.
[6] Chris J. Budd, Andrew T.T. McRae, Colin J. Cotter, The scaling and skewness of optimally transported meshes on the sphere, J. Comput. Phys. 375 (December 2018) 540–564.
[7] L. Chacón, G.L. Delzanno, J.M. Finn, Robust, multidimensional mesh-motion based on Monge-Kantorovich equidistribution, J. Comput. Phys. 230 (2011) 87–103.
[8] Chong Chen, Öktem Ozan, Indirect image registration with large diffeomorphic deformations, SIAM J. Imaging Sci. 11 (1) (2017).
[9] C.M. Duarte, A.M. Pyle, Stepping through an RNA structure: A novel approach to conformational analysis, J. Mol. Biol. 284 (5) (1998) 1465–1478.
[10] G.R. Ducharme, C. Vincent, C. Aliaume, A statistical test to detect vortices in the current fields of bodies of water, Environ. Ecol. Stat. 19 (3) (2012) 345–367.
[11] A. Figalli, L. Rifford, C. Villani, On the Ma-Trudinger-Wang curvature on surfaces, Calc. Var. 39 (2010) 307–332.
[12] B.D. Froese, Meshfree finite difference approximations for functions of the eigenvalues of the Hessian, Numer. Math. 138 (1) (2018) 75–99.
[13] V. Gorbunova, J. Sporring, P. Lo, M. Loeve, H.A. Tiddens, M. Nielson, A. Dirksen, M. de Bruijne, Mass preserving image registration for lung CT, Med. Image Anal. 16 (2012) 786–795.
[14] S. Haker, L. Zhu, A. Tannenbaum, S. Angenent, Optimal mass transport for registration and warping, Int. J. Comput. Vis. 60 (3) (2004) 225–240.
[15] B.F. Hamfeldt, A.G.R. Turnquist, A convergence framework for optimal transport on the sphere, arXiv preprint, arXiv:2103.05739, 2021.
[16] Brittany Froese Hamfeldt, Axel G.R. Turnquist, A convergent finite difference method for optimal transport on the sphere, J. Comput. Phys. 445 (November 2021).
[17] Brittany Froese Hamfeldt, Axel G.R. Turnquist, On the reduction in accuracy of finite difference schemes on manifolds without boundary, arXiv preprint, arXiv:2204.01892, 2022.
[18] J. Hempel, 3-Manifolds, American Mathematical Society, 1976.
[19] G. Loeper, On the regularity of solutions of optimal transportation problems, Acta Math. 202 (2009) 241–283.
[20] G. Loeper, Regularity of optimal maps on the sphere: The quadratic cost and the reflector antenna, Arch. Ration. Mech. Anal. 199 (1) (2011) 269–289.
[21] G. Loeper, C. Villani, Regularity of optimal transport in curved geometry: the nonfocal case, Duke Math. J. 151 (3) (February 2010) 431–485.
[22] X.-N. Ma, N.X. Trudinger, X.-J. Wang, Regularity of potential functions of the optimal transportation problem, Arch. Ration. Mech. Anal. 177 (2) (2005) 151–183.
[23] R.J. McCann, Polar factorization of maps on Riemannian manifolds, Geom. Funct. Anal. 11 (2001) 589–608.
[24] A.T. McRae, C.J. Cotter, C.J. Budd, Optimal-transport-based mesh adaptivity on the plane and sphere using finite elements, SIAM J. Sci. Comput. 40 (2) (2018) A1121–A1148.
[25] J.W. Milnor, On manifolds homeomorphic to the 7-sphere, Ann. Math. 64 (1956) 399–405.
[26] Klas Modin, Generalised Hunter-Saxton equations, optimal information transport, and factorisation of diffeomorphisms, J. Geom. Anal. 25 (2015) 1306–1334.
[27] T.A.E. Moselhy, Y.M. Marzouk, Bayesian inference with optimal maps, J. Comput. Phys. 231 (2012) 7815–7850.
[28] J. Moser, On the volume elements on a manifold, Trans. Am. Math. Soc. 120 (1965) 286–294.
[29] P. Niyogi, S. Smale, S. Weinberger, Finding the homology of submanifolds with high confidence from random samples, Discrete Comput. Geom. 39 (1–3) (2008) 419–441.
[30] A.M. Oberman, Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–Jacobi equations and free boundary problems, SIAM J. Numer. Anal. 44 (2) (2006) 879–895.

[31] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, J. Mol. Biol. 7 (1963) 95–99.
[32] Caleb Rottman, Martin Bauer, Klas Model, Sarang C. Joshi, Weighted diffeomorphic density matching with applications to thoracic image registration, in: 5th MICCAI Workshop on Mathematical Foundations of Computational Anatomy, 2015.
[33] C. Villani, Topics in Optimal Transportation, American Mathematical Society, 2003.
[34] C. Villani, Stability of a 4th-order curvature condition arising in optimal transport theory, J. Funct. Anal. 255 (2008) 2683–2708.
[35] Cédric Villani, Optimal Transport: Old and New, A Series of Comprehensive Studies in Mathematics, vol. 338, Springer, 2009.
[36] X.-J. Wang, On the design of a reflector antenna, IOP Sci. 12 (1996) 351–375.
[37] X.-J. Wang, On the design of a reflector antenna II, Calc. Var. Partial Differ. Equ. 20 (3) (2004) 329–341.
[38] Hilary Weller, Philip Browne, Chris Budd, Mike Cullen, Mesh generation on the sphere using Optimal Transport and the numerical solution of a Monge-Ampère type equation, J. Comput. Phys. 308 (March 2016) 102–123.
[39] L. Younes, Shapes and Diffeomorphisms, Applied Mathematical Sciences, vol. 171, Springer-Verlag, Berlin, 2010.
[40] Jiajia Yu, Rongjie Lai, Wuchen Li, Stanley Osher, Computational mean-field games on manifolds, arXiv preprint, arXiv:2206.01622, 2022.