A convergent finite difference method for optimal transport on the sphere

Brittany Froese Hamfeldt [*],[1], Axel G.R. Turnquist [2]

New Jersey Institute of Technology, 323 Martin Luther King Jr. Blvd., Newark, 07102, NJ, USA

ARTICLE INFO

ABSTRACT

We introduce a convergent finite difference method for solving the optimal transportation problem on the sphere. The method applies to both the traditional squared geodesic cost (arising in mesh generation) and a logarithmic cost (arising in the reflector antenna design problem). At each point on the sphere, we replace the surface PDE with a Generated Jacobian equation posed on the local tangent plane using geodesic normal coordinates. The discretization is inspired by recent monotone methods for the Monge-Ampère equation, but requires significant adaptations in order to correctly handle the mix of gradient and Hessian terms appearing inside the nonlinear determinant operator, as well as the singular logarithmic cost function. Numerical results demonstrate the success of this method on a wide range of challenging problems involving both the squared geodesic and the logarithmic cost functions.

© 2021 Elsevier Inc. All rights reserved.

We consider the problem of optimal transportation on the sphere. That is, given two prescribed density functions f_1 and f_2, we seek a mapping $T : \mathbb{S}^2 \to \mathbb{S}^2$ such that

$$T = \operatorname*{argmin}_{T_{\#} f_1 = f_2} \int_{\mathbb{S}^2} c(x, T(x)) f_1(x) dS(x). \tag{1}$$

Here $c(x, y)$ is the cost of transporting a unit of mass from x to y and $T_{\#} f_1 = f_2$ indicates that

$$\int_A f_1(x) \, dS(x) = \int_{T(A)} f_2(y) \, dS(y)$$

for every measurable $A \subset \mathbb{S}^2$.

In recent years, optimal transport has emerged as an important component in many different applications including image registration [1], astrophysics (estimating the shape of the early universe) [2], meteorology [3], machine learning [4], and geophysical inversion problems [5], among others. This has led to the development of a number of different numerical methods and convergence results for solving the optimal transport problem in Euclidean geometries [6–14].

---

* Corresponding author.
  E-mail addresses: bdfroese@njit.edu (B.F. Hamfeldt), agt6@njit.edu (A.G.R. Turnquist).

A less well studied but equally important setting is optimal transport between density functions on the sphere. Perhaps the simplest cost is the squared geodesic distance

$$c(x, y) = \frac{1}{2} d_{\mathbb{S}^2}(x, y)^2,$$

where $d_{\mathbb{S}^2}(x, y)$ denotes the geodesic distance between $x, y \in \mathbb{S}^2$. This cost function has recently been applied to the problem of mesh generation on the sphere in the context of meteorology [15,16].

A second cost of particular interest is the log cost

$$c(x, y) = -\log \|x - y\|,$$

which arises in the reflector antenna design problem [17,18]. The notation $\|\cdot\|$ denotes the Euclidean distance in the ambient space $\mathbb{R}^3$.

Recently, some progress has been made in the numerical solution of the Optimal Transport problem on the sphere in the case of the squared geodesic cost. The work of [16] used a geometric interpretation of a Monge-Ampère type equation on the sphere to produce the first such method. A finite element solution of this Monge-Ampère type equation was produced in [15]. For problems posed on a subset of the sphere, the stereographic projection can be used to reframe the problem as an optimal transport problem on the plane (with non-quadratic cost); this was the approach of [19]. The authors of the present article recently introduced a simple framework for proving the convergence of numerical methods for optimal transport on the sphere [20]. However, to date we are not aware of any methods that fit within this framework.

In this article, we produce the first convergent PDE-based method for solving the Optimal Transport problem on the sphere. This method is based on an approximation of a Generated Jacobian equation on local tangent planes, using carefully constructed local coordinates. The discretization is inspired by recent monotone methods that have been proposed for the Monge-Ampère equation. However, a complicating factor is the presence of gradient terms mixed with the Hessian terms inside of a nonlinear operator. This requires the introduction of new techniques for approximating both first- and second-order terms in order to preserve both the consistency and the monotonicity of our scheme. Additionally, the logarithmic cost function requires a careful regularization in order to preserve the well-posedness of the method. We produce an implementation and present computational results that demonstrate the success of this method for both the squared geodesic cost and the logarithmic cost.

1. Background

1.1. Optimal transport on the sphere

We consider points x, y lying on a unit sphere $\mathbb{S}^2$ centered at the origin. We are interested in two different cost functions $c(x, y)$: the squared geodesic distance on the sphere,

$$c(x, y) = \frac{1}{2} d_{\mathbb{S}^2}(x, y)^2 = \frac{1}{2}\left(2\sin^{-1}\left(\frac{\|x - y\|}{2}\right)\right)^2, \tag{2}$$

and the log-cost arising in the reflector antenna problem,

$$c(x, y) = -\log\|x - y\|. \tag{3}$$

The optimal map corresponding to each cost function is determined from the conditions

$$\begin{cases} \nabla_{\mathbb{S}^2, x} c(x, T(x, p)) = -p, & x \in \mathbb{S}^2, p \in \mathcal{T}_x \\ T(x, p) \in \mathbb{S}^2 \end{cases} \tag{4}$$

where $\mathcal{T}_x$ denotes the tangent plane at x. These mappings can be found explicitly for the cost functions we are interested in. In the case of the squared geodesic cost [15], the map is

$$T(x, p) = \cos(\|p\|) x + \sin(\|p\|) \frac{p}{\|p\|}. \tag{5}$$

In the case of the logarithmic cost [20], the map is

$$T(x, p) = x \frac{\|p\|^2 - 1/4}{\|p\|^2 + 1/4} - \frac{p}{\|p\|^2 + 1/4}. \tag{6}$$

The solution to the optimal transport problem then satisfies

$$F\left(x, \nabla_{\mathbb{S}^2} u(x), D^2_{\mathbb{S}^2} u(x)\right) = 0 \tag{7}$$

where

$$F(x, p, M) \equiv -\det(M + A(x, p)) + H(x, p). \tag{8}$$

Here

$$A(x, p) = D^2_{\mathbb{S}^2, xx} c(x, T(x, p))$$

$$H(x, p) = \left| \det D^2_{\mathbb{S}^2, xy} c(x, T(x, p)) \right| f_1(x) / f_2(T(x, p)),$$

and the PDE now describes a nonlinear relationship between the surface gradient and Hessian on the sphere.

In order to ensure existence of a unique solution that coincides with the correct solution of the optimal transport problem, it is necessary to require the solution u to satisfy a generalized notion of convexity known as c-convexity.

Definition 1 (*c-convexity*). A function $u : \mathbb{S}^2 \to \mathbb{R}$ is c-convex if for every $x_0 \in \mathbb{S}^2$ there exists some $y_0 \in \mathbb{S}^2$ such that

$$u(x) \geq u(x_0) + c(x_0, y_0) - c(x, y_0)$$

for every $x \in \mathbb{S}^2$.

This definition indicates that a c-convex function u must everywhere be supported by a function whose shape is determined by the negative cost function $-c(x, y_0)$. We remark that in Euclidean space with the cost $c(x, y) = -x \cdot y$, this recovers the usual notion that a convex function should everywhere have a supporting hyperplane.

The second-order optimality conditions corresponding to Definition 1 leads to the following constraint, which is sufficient to ensure ellipticity (and well-posedness) of the PDE (8).

$$D^2_{\mathbb{S}^2} u(x) + A(x, \nabla_{\mathbb{S}^2} u(x)) \geq 0. \tag{9}$$

The PDE (8) belongs to the class of Generated Jacobian equations [21], and places constraints on the Jacobian of the mapping $T(x, \nabla u(x))$ in order to force the density f_1 to be transported into the density f_2. It is similar to the Monge-Ampère equation that is seen in optimal transport in Euclidean space (with quadratic cost $c(x, y) = \frac{1}{2}\|x - y\|^2$). However, the introduction of more complicated geometries and cost functions now leads to a mix of gradient and Hessian terms inside of the nonlinear determinant operator.

This problem was studied by Loeper [22], who showed that under mild regularity requirements on the data, the optimal transport problem on the sphere admits a smooth (C^3) solution u.

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

(a) There exists some $m > 0$ such that $f_2(x) \geq m$ for all $x \in \mathbb{S}^2$.
(b) The mass balance condition holds, $\int_{\mathbb{S}^2} f_1(x)\, dx = \int_{\mathbb{S}^2} f_2(y)\, dy$.
(c) The cost function is either $c(x, y) = \frac{1}{2} d_{\mathbb{S}^2}(x, y)^2$ or $c(x, y) = -\log\|x - y\|$.
(d) The data satisfies the regularity requirements $f_1, f_2 \in C^{1,1}(\mathbb{S}^2)$.

Weak (C^1) solutions are also possible for discontinuous density functions that are only in L^p ($1 \leq p < \infty$). Moreover, the solution is unique up to additive constants.

1.2. Approximation of elliptic PDEs

The PDE operators we consider in this work are degenerate elliptic.

$$F\left(x, u(x), \nabla u(x), D^2 u(x)\right) = 0, \quad x \in \mathbb{S}^2. \tag{10}$$

Definition 3 (*Degenerate elliptic*). The operator $F : \mathbb{S}^2 \times \mathbb{R} \times \mathcal{S}^2 \to \mathbb{R}$ is *degenerate elliptic* if

$$F(x, u, X) \leq F(x, v, Y)$$

whenever $u \leq v$ and $X \geq Y$.

The PDE operators that we consider in this work are degenerate elliptic if they are non-decreasing functions of the argument u and non-increasing functions of all subsequent arguments (which involve second directional derivatives).

Since degenerate elliptic equations need not have classical solutions, solutions may need to be interpreted in a weak sense. Moreover, even when smooth solutions exist, the use of weak solutions often provides an easier path to convergent numerical methods for fully nonlinear elliptic equations. The numerical methods developed in this article are guided by the very powerful concept of the viscosity solution [23].

Definition 4 *(Upper and lower semi-continuous envelopes).* The *upper and lower semi-continuous envelopes* of a function $u(x)$ are defined, respectively, by

$$u^*(x) = \limsup_{\mathbf{y} \to x} u(\mathbf{y}), \quad u_*(x) = \liminf_{\mathbf{y} \to x} u(\mathbf{y}).$$

Definition 5 *(Viscosity solution).* An upper (lower) semi-continuous function u is a *viscosity subsolution (supersolution)* of (10) if for every $\phi \in C^2(\bar{\Omega})$, whenever $u - \phi$ has a local maximum (minimum) at $x \in \bar{\Omega}$, then

$$F_*^{(*)}(x, u(x), \nabla\phi(x), D^2\phi(x)) \leq (\geq)0.$$

A continuous function u is a *viscosity solution* of (10) if it is both a subsolution and a supersolution.

The first steps towards constructing convergent methods for fully nonlinear elliptic equations where provided by a powerful framework introduced by Barles and Souganidis [24] and further developed by Oberman [25].

We consider finite difference schemes that have the form

$$F^h(x, u(x), u(x) - u(\cdot)) = 0 \tag{11}$$

where h is a small parameter relating to the grid resolution. The value of these approximation schemes depends on the spatial location $x \in \mathbb{S}^2$, the value of the function u at this reference point x, and differences of the form $u(x) - u(y)$ where y is chosen from one or more other points in the domain. The notation $u(\cdot)$ is used to indicate this dependence on the values of u at one or more different points in the domain.

The convergence framework requires notions of consistency, monotonicity and stability.

Definition 6 *(Consistency).* The scheme (11) is *consistent* with the equation (10) if for any smooth function ϕ and $x \in \bar{\Omega}$,

$$\limsup_{h \to 0^+, \mathbf{y} \to x, \xi \to 0} F^h(\mathbf{y}, \phi(\mathbf{y}) + \xi, \phi(\mathbf{y}) - \phi(\cdot)) \leq F^*(x, \phi(x), \nabla\phi(x), D^2\phi(x)),$$

$$\liminf_{h \to 0^+, \mathbf{y} \to x, \xi \to 0} F^h(\mathbf{y}, \phi(\mathbf{y}) + \xi, \phi(\mathbf{y}) - \phi(\cdot)) \geq F_*(x, \phi(x), \nabla\phi(x), D^2\phi(x)).$$

Definition 7 *(Monotonicity).* The scheme (11) is monotone if F^h is a non-decreasing function of its final two arguments.

Definition 8 *(Stability).* The scheme (11) is stable if there exists some $M \in \mathbb{R}$ such that if u^h is any solution of (11) then $\|u^h\|_\infty \leq M$.

To consistent schemes, we can also associate a local truncation error.

Definition 9 *(Truncation error).* The truncation error $\tau(h) > 0$ of the scheme (11) is a quantity chosen so that for every smooth function ϕ

$$\limsup_{h \to 0} \max_{x \in \mathcal{G}^h} \frac{\left| F^h(x, \phi(x), \phi(x) - \phi(\cdot)) - F(x, \nabla\phi(x), D^2\phi(x)) \right|}{\tau(h)} < \infty.$$

Schemes that satisfy these three properties respect the notion of the viscosity solution at the discrete level. In particular, these schemes preserve the maximum principle and are guaranteed to converge to the solution of the underlying PDE if that equation satisfies a comparison principle.

Definition 10 *(Comparison principle).* A PDE has a *comparison principle* if whenever u is an upper semi-continuous subsolution and v a lower semi-continuous supersolution of the equation, then $u \leq v$ on $\bar{\Omega}$.

Theorem 11 *(Convergence [25]). Let u be the unique viscosity solution of the PDE (10), where F is a degenerate elliptic operator with a comparison principle. Let the approximation F^h be consistent, monotone, and stable and u^h any solution of the scheme (11). Then u^h converges uniformly to u as $h \to 0$.*

1.3. Convergence results for optimal transport on the sphere

The nonlinear PDE (8) that we are interested in does not satisfy a comparison principle, and thus the Barles-Souganidis convergence results do not immediately apply. However, the authors of this article have recently demonstrated that this framework could be extended to (8) through a careful consideration of the geometry and the introduction of a term that enforces a strong form of stability [20].

1.3.1. Reformulation of PDE

The first step in constructing convergent methods is to translate the surface PDE (8) at the point $x_0 \in \mathbb{S}^2$ to an equation posed on the local tangent plane $\mathcal{T}_{x_0}$. This requires introducing local coordinates on the tangent plane. In general, local coordinates will introduce distortions to the Hessian that require the introduction of additional gradient terms. However, this problem can be avoided with the use of *geodesic normal coordinates*, which preserve distance from the reference point x_0. An explicit expression for these coordinates $v_{x_0} : \mathbb{S}^2 \setminus \{-x_0\} \to \mathcal{T}_{x_0}$ is given by

$$v_{x_0}(x) = x_0 \left(1 - d_{\mathbb{S}^2}(x_0, x) \cot d_{\mathbb{S}^2}(x_0, x) \right) + x \left(d_{\mathbb{S}^2}(x_0, x) \csc d_{\mathbb{S}^2}(x_0, x) \right). \tag{12}$$

For each point $x_0 \in \mathbb{S}^2$ we can now define a function $\tilde{u}_{x_0}(z)$ on the relevant tangent plane $\mathcal{T}_{x_0}$ in a neighborhood of x_0 by

$$\tilde{u}_{x_0}(z) = u \left(v_{x_0}^{-1}(z) \right). \tag{13}$$

This choice of coordinates allows us to express the PDE (8) at the point $x_0 \in \mathbb{S}^2$ as a generalized Monge-Ampère equation

$$F(x_0, \nabla \tilde{u}(x_0), D^2 \tilde{u}(x_0)) \equiv -\det \left(D^2 \tilde{u}(x_0) + A(x_0, \nabla \tilde{u}(x_0)) \right) + H(x_0, \nabla \tilde{u}(x_0)) = 0, \tag{14}$$

where all derivatives are now interpreted in the usual sense on a two-dimensional plane.

Remark 12. From this point on, we will refer to the surface PDE (8) and the tangent plane representation (14) interchangeably, with the implicit understanding that coordinates in the tangent plane are given by the geodesic local coordinates.

The problem of approximating the Optimal Transport PDE on the sphere can be further simplified by embedding the c-convexity (ellipticity) constraint into the equation. This is accomplished through the introduction of a modified determinant operator satisfying

$$\det{}^+(M) = \begin{cases} \det(M), & M \geq 0 \\ \leq 0, & \text{otherwise.} \end{cases} \tag{15}$$

Then we can absorb the constraint into the PDE (14) through the modification

$$F^+(x, \nabla u(x), D^2 u(x)) \equiv -\det{}^+(D^2 u(x) + A(x, \nabla u(x))) + H(x, \nabla u(x)) = 0. \tag{16}$$

Finally, the solutions of the Optimal Transport PDE satisfy *a priori* Lipschitz bounds $\|\nabla u\| < R$ for any $R > \pi$ (squared geodesic cost) or $R > C$ (logarithmic cost; see [22, Proposition 6.1] for details). These bounds can be explicitly built into the PDE through a further modification

$$G(x, \nabla u(x), D^2 u(x)) \equiv \max \left\{ F^+(x, \nabla u(x), D^2 u(x)), \|\nabla u(x)\| - R \right\} = 0. \tag{17}$$

While this modification enforces a condition that is automatically satisfied by solutions at the continuous level, it also improves the stability of approximation schemes.

1.3.2. Convergence results

The convergence results of [20] require a mesh or point cloud $\mathcal{G}^h \subset \mathbb{S}^2$ on the sphere that satisfies very mild structural regularity conditions. We define the discretization parameter h as

$$h = \sup_{x \in \mathbb{S}^2} \min_{y \in \mathcal{G}^h} d_{\mathbb{S}^2}(x, y). \tag{18}$$

In particular, this guarantees that any ball of radius h on the sphere will contain at least one discretization point.

Then we require a grid that admits a triangulation without any long, skinny triangles. Specifically:

Hypothesis 13 (*Conditions on point cloud*). There exists a triangulation T^h of $\mathcal{G}^h$ with the following properties:

(a) The diameter of the triangulation, defined as

$$\operatorname{diam}(T^h) = \max_{t \in T^h} \operatorname{diam}(t), \tag{19}$$

satisfies $\operatorname{diam}(T^h) \to 0$ as $h \to 0$.

(b) There exists some $\gamma < \pi$ (independent of h) such that whenever θ is an interior angle of any triangle $t \in T^h$ then $\theta \leq \gamma$.
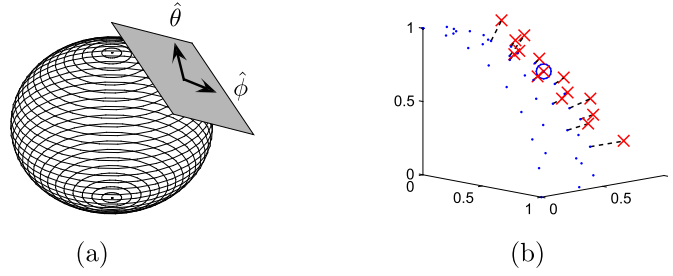
Fig. 1. (a) The sphere $\mathbb{S}^2$ and tangent plane $\mathcal{T}_{x_0}$. (b) A point cloud discretizing one octant of the unit sphere ($\cdot$), the point x_0 (o), and the projections z of neighboring nodes onto $\mathcal{T}_{x_0}$ ($\times$).

We remark that there is no need to construct this triangulation in practice, it need only exist in theory.

We also associate to each point cloud $\mathcal{G}^h$ a search radius $r(h)$ chosen to satisfy

$$r(h) \to 0, \quad \frac{h}{r(h)} \to 0 \text{ as } h \to 0, \quad \operatorname{diam}(T^h) < r(h). \tag{20}$$

Next, we project nearby grid points onto the local tangent plan $\mathcal{T}_{x_0}$, which is spanned by the orthonormal vectors $\left(\hat{\theta}, \hat{\phi}\right)$. For all points $x_i \in \mathcal{G}^h \cap B(x_0, r(h))$, we define their projection onto the tangent plane through geodesic normal coordinates via

$$z_i = x_0 \left(1 - d_{\mathbb{S}^2}(x_0, x_i) \cot d_{\mathbb{S}^2}(x_0, x_i)\right) + x_i \left(d_{\mathbb{S}^2}(x_0, x_i) \csc d_{\mathbb{S}^2}(x_0, x_i)\right). \tag{21}$$

Let $\mathcal{Z}^h(x_0) \subset \mathcal{T}_{x_0}$ be the resulting collection of points. See Fig. 1.

These are now the discretization points available to use for the approximation of (17) at x_0; recall that this PDE is posed on the two-dimensional tangent plane. The convergence result of [20] proposed a specific, simple discretization E^h of the Eikonal term $\|\nabla u(x_0)\|$ that is valid on any grid or point cloud.

$$E^h(z, u(z) - u(\cdot)) = \max_{y \in \mathcal{Z}^h(z)} \frac{u(z) - u(y)}{\|z - y\|}. \tag{22}$$

Then letting F^h be an approximation of the convexified PDE (16), we can define an approximation of the modified PDE (17) by

$$G^h(z, u(z) - u(\cdot)) = \max \left\{ F^h(z, u(z) - u(\cdot)), E^h(z, u(z) - u(\cdot)) - R \right\}, \quad z \in \mathcal{G}^h. \tag{23}$$

Fixing a point $x^* \in \mathbb{S}^2$ and a sequence $x_h^* \in \mathcal{G}^h \to x^*$, we arrive at the following two-step approach for obtaining a numerical solution u^h of the Optimal Transport problem on the sphere.

1. Solve the discrete system

$$G^h \left(z, v^h(z) - v^h(\cdot)\right) + \tau(h) v^h(z) = 0, \quad z \in \mathcal{G}^h \tag{24}$$

for the grid function v^h.

2. Define the candidate solution

$$u^h(z) = v^h(z) - v^h(x_h^*), \quad z \in \mathcal{G}^h. \tag{25}$$

Under appropriate assumptions on the density functions f_1, f_2 and the approximation scheme F^h, this is guaranteed to converge to the solution of the Optimal Transport problem on the sphere.

Theorem 14 (*Convergence of approximation [20]*). *Under the assumptions of Hypothesis 2, let $u \in C^3(\mathbb{S}^2)$ be the unique solution of (8) satisfying $u(x^*) = 0$. Let $\mathcal{G}^h$ be a grid satisfying Hypothesis 13 and let F^h be any consistent, monotone approximation of (17). Then for each sufficiently small $h > 0$, the grid function u^h defined in (25) is uniquely defined. Moreover, u^h converges uniformly to u as $h \to 0$.*

Remark 15. The above convergence theorem can be extended to weak (C^1) solutions by additionally requiring the approximation scheme to underestimate the value of the PDE.

2. Formulation of the PDE

The PDE (8)-(9) can be formulated in different equivalent ways. We begin by briefly describing a reformulation that lends itself to the construction of a discretization that fits within the framework of Theorem 14.

2.1. Regularization of logarithmic cost

One modification of the PDE that we find is necessary to build monotone schemes is to make the logarithmic cost Lipschitz by using a cutoff function. We recall (as discussed in subsubsection 1.3.1) that the solution u to the optimal transport satisfies an *a priori* Lipschitz bound [22] $\|\nabla u\| < R$, which also yields the following lower bound on the distance mass can be transported.

$$\|x - y\| > \frac{1}{\sqrt{R^2 + 1/4}} \tag{26}$$

when $y = T(x, \nabla u(x))$ is the exact transport map.

However, the process of solving a discrete version of (8) may evolve through values of u (and consequently y) that do not satisfy this bound. This loss of Lipschitz continuity can lead to a breakdown in monotonicity. We thus propose the following C^3 regularization of the logarithmic cost function, which agrees with the true logarithmic cost when the bound (26) is satisfied.

$$\tilde{c}(x, y) = \begin{cases} -\log \|x - y\|, & \|x - y\| \geq \frac{1}{\sqrt{R^2 + 1/4}} \\ \Psi(\|x - y\|), & \|x - y\| < \frac{1}{\sqrt{R^2 + 1/4}} \end{cases} \tag{27}$$

where $z_* = \dfrac{1}{\sqrt{R^2 + 1/4}}$ and

$$\Psi(z) = -\log(z_*) - \frac{1}{z_*}(z - z_*) + \frac{1}{2z_*^2}(z - z_*)^2 - \frac{1}{3z_*^3}(z - z_*)^3.$$

Because this regularization does not change the solutions of the PDE, the analysis of [20] and the ultimate convergence result (Theorem 14) will also apply to discretizations involving this smoother cost function.

Lemma 16 (*Equivalence of solution for modified cost function*). *Under the assumptions of Hypothesis 2, a function $u \in C^3(\mathbb{S}^2)$ is a solution of (8) with the logarithmic cost (3) if and only if it is a solution of (8) with the regularized cost (27).*

Proof. First let u be a solution using the original cost function $c(x, y) = -\log \|x - y\|$. Because of the *a priori* bounds on the gradient, this automatically satisfies the PDE (8) with the regularized cost function.

Next let v be any solution of (8) using the regularized cost function $\tilde{c}(x, y)$ from (27). Notice that for $\|x - y\| \geq \frac{1}{\sqrt{R^2 + 1/4}}$, we have

$$\tilde{c}(x, y) = -\log\left(2\sin\left(\frac{1}{2}d_{\mathbb{S}^2}(x, y)\right)\right). \tag{28}$$

It is easily verified via differentiation that this is convex in $d_{\mathbb{S}^2}(x, y)$ for $\|x - y\| < \dfrac{1}{\sqrt{R^2 + 1/4}}$, and the original logarithmic cost is also convex in $d_{\mathbb{S}^2}(x, y)$. The modified cost is C^3, so verifying that the second derivative in the variable $d_{\mathbb{S}^2}(x, y)$ is everywhere positive is sufficient to guarantee convexity in this variable. Since this new cost function is Lipschitz and convex in the Riemannian distance, the $\tilde{c}$-convex solution of (8) for the regularized cost $\tilde{c}(x, y)$ is guaranteed to be unique by McCann as noted in [26]. Since the solution u of the original equation solves this modified equation, uniqueness requires that $v = u$. $\square$

Remark 17. Because this regularization transforms the Optimal Transport problem with a singular cost function into an Optimal Transport problem with a smooth cost function, the techniques of [20] (which were introduced for the squared geodesic cost) can be extended to show that weak ($C^{0,1}$) solutions of this modified problem are also unique. This assumption of uniqueness of $C^{0,1}$ viscosity solutions was needed to prove convergence even in the smooth setting. With minor modification, Theorem 14 can also be extended to handle convergence to weak solutions under much milder regularity requirements on the data; see [20, Theorem 36]. These modifications are addressed in subsection 4.5.

In the development and analysis of our numerical method in the following sections, we will often refer to the cost function $\tilde{c}$. In the case of the logarithmic cost, this refers to the regularized cost (27). In the case of the squared geodesic cost, $\tilde{c}$ will refer to the original cost function c, which is automatically smooth.

2.2. Variational formulation of the determinant of a Hessian

Since our PDE involves computing the determinant of a Hessian, here we show how to do this so as to later build a monotone discretization of the second derivatives. As utilized in [27], Hadamard's inequality allows the determinant of a positive definite matrix M to be computed via the minimization problem

$$\det M = \min_{v_i \in V} \prod_{i=1}^{d} v_i^T M v_i \tag{29}$$

where V is the set of all orthogonal bases for $\mathbb{R}^d$.

We require a formulation satisfying (15), which modifies this formulation to ensure that no negative terms appear when the symmetric matrix M is not positive definite. A simple approach is to use

$$\det^+ M = \min_{v_i \in V} \prod_{i=1}^{d} \max \left\{ v_i^T M v_i, 0 \right\}. \tag{30}$$

If our matrix $M = D^2 \phi(x)$ is a Hessian matrix, this becomes:

$$\det^+ (D^2 \phi(x)) = \min_{v_1, v_2 \in V} \left\{ \prod_{j=1}^{2} v_j^T (D^2 \phi(x)) v_j, 0 \right\}$$

$$= \min_{v_1, v_2 \in V} \prod_{j=1}^{2} \max \left\{ \frac{\partial^2 \phi}{\partial v_j^2}, 0 \right\}.$$

In particular, this allows us to replace the determinant in (8) with

$$\det^+ (D^2 u(x) + A(x, \nabla u(x))) = \min_{v_1, v_2 \in V} \prod_{j=1}^{2} \max \left\{ \frac{\partial^2 u(x)}{\partial v_j^2} + \frac{\partial^2 c(x, y)}{\partial v_j^2} \bigg|_{y = T(x, \nabla u(x))}, 0 \right\}. \tag{31}$$

2.3. Mixed Hessian

The framework developed by the authors of this article in [20] only requires the construction of consistent approximations of derivatives with respect to x, expressed in the geodesic normal coordinate system. For the mixed Hessian term, $\det D_{xy}^2 c(x, y)$ it is not immediately clear how to do this. In fact, the relative simplicity of the notation obfuscates the actual complexity of the object. In [26], Loeper outlines a clearer representation of this quantity in curved geometries and for different cost functions.

We recall that the optimal map $T(x, p)$ (also known as the c-exponential map) satisfies (4) and can be constructed explicitly for the cost function of interest to us via (5)-(6). Then from [26], the mixed Hessian satisfies

$$[D_{xy}^2 c]^{-1} = -D_p T(x, p)|_{x, p = -\nabla_x c(x, y)}. \tag{32}$$

This representation formula (32) shows that the inverse of the mixed Hessian is simply the Jacobian of the map $T(x, p)$ with respect to p. Since we will be taking the determinant, this is actually a change of area formula for the transformation $T(x, p)$. See Fig. 2.

We will compute this change of area by computing the linear differential map dT for both cost functions. That is, the change in area will be computed using orthogonal perturbations Δp_1, Δp_2 (so $\Delta p_1 \cdot \Delta p_2 = 0$) in the tangent plane $\mathcal{T}_x$.

$$\left| \det \left(D_p T(x, p) \right) \right| =$$

$$\left| \lim_{\|\Delta p_1\| \to 0} \frac{T(x, p + \Delta p_1) - T(x, p)}{\|\Delta p_1\|} \times \lim_{\|\Delta p_2\| \to 0} \frac{T(x, p + \Delta p_2) - T(x, p)}{\|\Delta p_2\|} \right|.$$

In general, the area element on a manifold is a function of the wedge product of two covectors, which need not be orthogonal. This has the interpretation of the area of a parallelogram on the manifold. However, in the special case where the vectors are orthogonal (or are orthogonal to leading order), this reduces to an ordinary product. This is indeed the case for both the squared geodesic and logarithmic cost functions. Thus the change of area formula reduces to the simpler expression
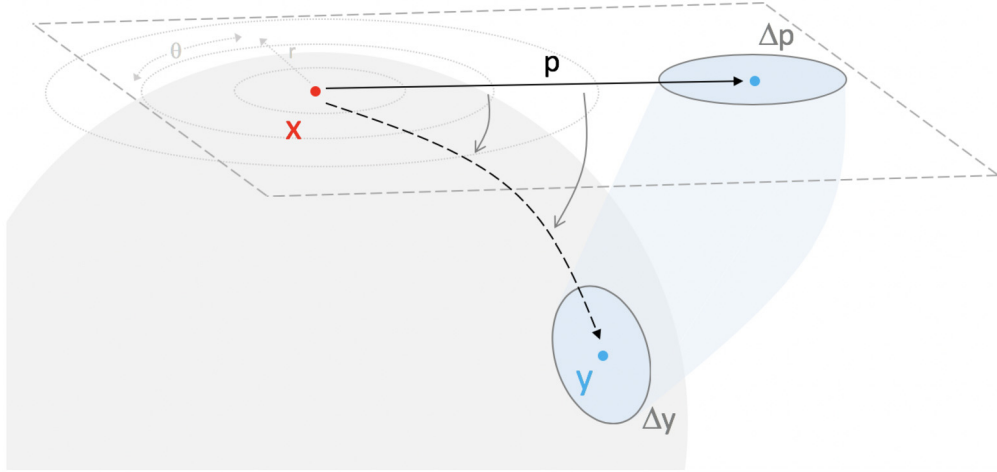
Fig. 2. The determinant of the mixed Hessian is the change in area formula from the set $T(x, E)$ on the sphere to the set E on the local tangent plane.

$$\left| \det\left(D_p T(x, p) \right) \right| =$$

$$\lim_{\|\Delta p_1\|, \|\Delta p_2\| \to 0, \, \Delta p_1 \cdot \Delta p_2 = 0} \frac{d_{\mathbb{S}^2}\left(T(x, p), T(x, p + \Delta p_1) \right) d_{\mathbb{S}^2}\left(T(x, p), T(x, p + \Delta p_2) \right)}{\|\Delta p_1\| \, \|\Delta p_2\|}$$

and the determinant of the mixed Hessian is given by

$$\left| \det(D_{xy}^2 c(x, y)) \right| = \left. \frac{1}{\left| D_p T(x, p) \right|} \right|_{p = -\nabla_x c(x, y)}. \tag{33}$$

This can be computed explicitly for both the squared geodesic cost,

$$\left| \det(D_{xy}^2 c(x, y)) \right| = \frac{\|p\|}{\sin \|p\|}, \tag{34}$$

and for the logarithmic cost,

$$\left| \det(D_{xy}^2 c(x, y)) \right| = \left(\|p\|^2 + 1/4 \right)^2. \tag{35}$$

See Appendix A for details.

We note also that these formulas coincide with the formulas that can be derived via standard change of variables formulas by requiring

$$\int_{T(x,E)} dS = \int_E \left| \det(D_p T(x, p)) \right| dp$$

for every measurable $E \in \mathcal{T}(x)$.

3. Numerical method

We now explain how we actually construct an approximation scheme for (8) at a point $x \in \mathbb{S}^2$. In section 4, we will demonstrate that this method does converge to the true solution of the optimal transport problem.

3.1. Construction of finite difference stencils

We begin with a point cloud $\mathcal{G} \in \mathbb{S}^2$ that discretizes the sphere; we assume only the minimal regularity required by Hypothesis 13.

We begin by considering a fixed point $x_i \in \mathcal{G}$ and establishing a computational neighborhood $N(i)$ about this point. For any fixed $C > 0$, we define

$$N(i) = \{ j \mid x_j \in \mathcal{G}, \, d_{\mathbb{S}^2}(x_i, x_j) \le C\sqrt{h} \}. \tag{36}$$

Once $N(i)$ is established, the points $x_j \in N(i)$ are projected on to the local tangent plane $\mathcal{T}_{x_i}$ via a geodesic normal coordinate projection (12).
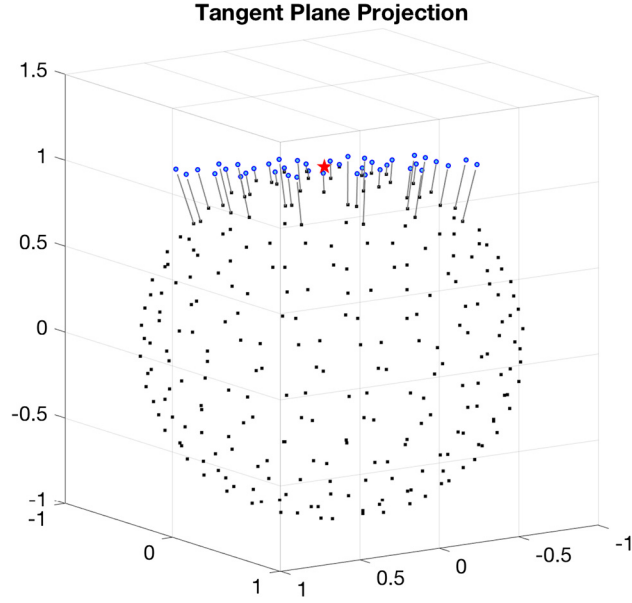
Fig. 3. Projection onto the tangent plane via geodesic normal coordinates. The points in $\mathcal{G}$ are indicated with small black squares, the computational point x_i is indicated as a red star, and the projection of the neighborhood $N(i)$ to the local tangent plane is indicated with blue circles. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)
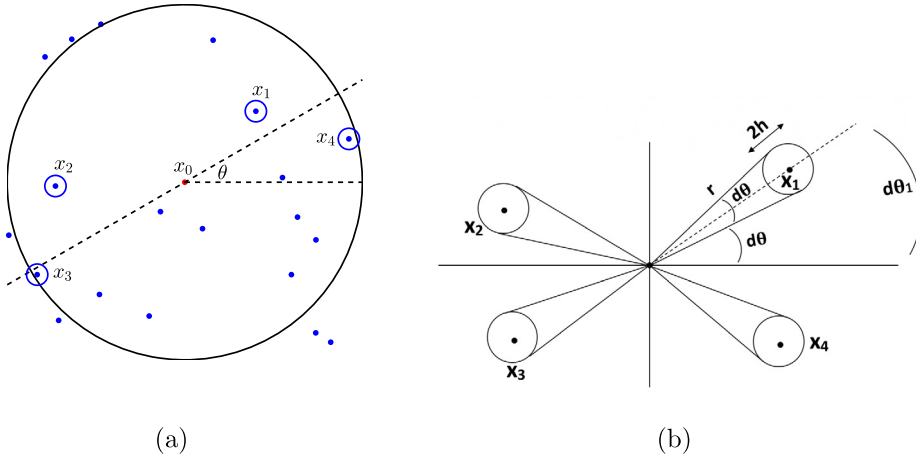

Fig. 4. Choice of computational points $x_{v,j}$ in the local tangent plane.

$$z_j = v_{x_i}(x_j).$$

We denote the resulting point cloud on the tangent plane by

$$\mathcal{N}(i) = \{z_j \mid j \in N(i)\} \subset \mathcal{T}_{x_i}.$$

Fig. 3 shows an exaggerated example of this tangent plane projection.

Next we suppose that we are interested in resolving behavior along some direction $v \in \mathbb{R}^2$. Following a slight modification of [28], we select four points $x_{v,j} \in \mathcal{N}(i)$ that are well-aligned with the direction v. See Fig. 4(a).

We introduce the following notation:

- $r = C\sqrt{h}$ is the search radius used to define the neighborhood $N(i)$. For all $x_{v,j} \in \mathcal{N}(i)$ we have that $\|x_{v,j} - x_i\| \le r$.
- $\theta_{v,j}$ is the angle between $x_{v,j} - x_i$ and the direction v.
- $d\theta_{v,j}$ is the minimal absolute angle between $x_{v,j} - x_i$ and the direction v
- $d\theta$ is the overall angular resolution of the stencil, which is related to the search radius through $d\theta = \dfrac{2h}{r} + \mathcal{O}(h) = \mathcal{O}(\sqrt{h})$; see [28].

- Each $x_{v,j}$ can be represented in polar coordinates as $(h_{v,j}, \theta_{v,j})$ using the coordinate system where x_i is the origin and the coordinate directions $v, v^\perp$ are orthogonal.
- Components of $x_{v,j}$ can be expressed using the shorthand notation

$$C_{v,j} = h_{v,j}\cos\theta_{v,j}, \quad S_{v,j} = h_{v,j}\sin\theta_{v,j}.$$

The following lemma follows immediately from the proof of [28, Lemma 11]. Fig. 4(b) illustrates the four small balls where each of these four neighbors is required to exist.

Lemma 18 (*Properties of neighbors*). *For every $x_i \in \mathcal{G}$ and $v \in \mathbb{R}^2$, four neighbors $x_{v,j} \in \mathcal{N}(i)$ exist satisfying the following properties:*

- $x_{v,j}$ *resides in the jth quadrant.*
- *The angular component of $x_{v,j}$ satisfies $d\theta \leq d\theta_{v,j} \leq 2d\theta$.*
- *The radial component of $x_{v,j}$ satisfies $r - 2h \leq h_{v,j} \leq r$.*

These additional requirements on stencil will be critical to developing monotone approximations of functions of the gradient.

3.2. Approximation of second derivatives

Our overall approximation of (8) will hinge on the construction of (negative) monotone schemes for second directional derivatives $\dfrac{\partial^2 \phi}{\partial v^2}$. We introduce approximations of the form

$$\mathcal{D}_{vv}\phi(x_i) = \sum_{j=1}^{4} a_{v,j}\left(\phi(x_{v,j}) - \phi(x_i)\right). \tag{37}$$

As in [28], consistency and negative monotonicity can be achieved by finding a solution of the system

$$\begin{cases} \sum a_{v,j}h_{v,j}\cos\theta_{v,j} = 0 \\ \sum a_{v,j}h_{v,j}\sin\theta_{v,j} = 0 \\ \sum \frac{1}{2}a_{v,j}h_{v,j}^2\cos^2\theta_{v,j} = 1 \\ a_{v,j} \geq 0 \end{cases} \tag{38}$$

An explicit solution is given by

$$\begin{aligned} a_{v,1} &= \frac{2S_{v,4}(C_{v,3}S_{v,2} - C_{v,2}S_{v,3})}{\det(A)} \\ a_{v,2} &= \frac{2S_{v,3}(C_{v,1}S_{v,4} - C_{v,4}S_{v,1})}{\det(A)} \\ a_{v,3} &= \frac{-2S_{v,2}(C_{v,1}S_{v,4} - C_{v,4}S_{v,1})}{\det(A)} \\ a_{v,4} &= \frac{-2S_{v,1}(C_{v,3}S_{v,2} - C_{v,2}S_{v,3})}{\det(A)} \end{aligned} \tag{39}$$

where

$$\begin{aligned} \det(A) =& (C_{v,3}S_{v,2} - C_{v,2}S_{v,3})(C_{v,1}^2 S_{v,4} - C_{v,4}^2 S_{v,1}) \\ &- (C_{v,1}S_{v,4} - C_{v,4}S_{v,1})(C_{v,3}^2 S_{v,2} - C_{v,2}^2 S_{v,3}). \end{aligned} \tag{40}$$

3.3. Approximation of functions of the gradient

The PDE (8) involves several terms of the form $g(\nabla u)$. Importantly, either automatically or through appropriate regularization (see subsection 2.1), each of these functions g has a bounded Lipschitz constant L_g. This allows us to pursue a generalized Lax-Friedrichs type discretization of the form

$$\tilde{g}^{\pm}(\mathcal{D}\phi(x_i)) = g\left(\sum_{\nu \in \{(1,0),(0,1)\}} \nu \sum_{j=1}^{4} b_{\nu,j}\left(\phi(x_{\nu,j}) - \phi(x_i)\right)\right)$$

$$\mp \epsilon_g \sum_{\nu \in \{(1,0),(0,1)\}} \sum_{j=1}^{4} a_{\nu,j}\left(\phi(x_{\nu,j}) - \phi(x_i)\right). \tag{41}$$

Above, the coefficients $a_{\nu,j}$ are identical to the coefficients that arise in the approximation of second directional derivatives. This introduces a Laplacian regularization term, which is carefully chosen to enforce monotonicity (or negative monotonicity) even if the coefficients $b_{\nu,j}$ do not on their own produce a monotone scheme.

We first require coefficients $b_{\nu,j}$ that ensure that

$$\mathcal{D}_{\nu}\phi(x_i) = \sum_{j=1}^{4} b_{\nu,j}\left(\phi(x_{\nu,j}) - \phi(x_i)\right)$$

is a consistent approximation of the first directional derivative $\dfrac{\partial \phi(x_i)}{\partial \nu}$.

Taylor expanding, we obtain

$$\mathcal{D}_{\nu}\phi(x_i) = \sum_{j=1}^{4} b_{\nu,j}\left(h_{\nu,j}\cos\theta_{\nu,j}\frac{\partial\phi(x_i)}{\partial\nu} + h_{\nu,j}\sin\theta_{\nu,j}\frac{\partial\phi(x_i)}{\partial\nu^{\perp}} + \mathcal{O}(r^2)\right).$$

Consistency then requires a solution of the system

$$\begin{cases} \sum b_{\nu,j}h_{\nu,j}\cos\theta_{\nu,j} = 1 \\ \sum b_{\nu,j}h_{\nu,j}\sin\theta_{\nu,j} = 0. \end{cases} \tag{42}$$

An explicit solution is

$$b_{\nu,1} = \frac{S_{\nu,4}(S_{\nu,3}C_{\nu,2}^2 - S_{\nu,2}C_{\nu,3}^2)}{\det(A)}$$

$$b_{\nu,2} = -\frac{S_{\nu,3}(S_{\nu,4}C_{\nu,1}^2 - S_{\nu,1}C_{\nu,4}^2)}{\det(A)}$$

$$b_{\nu,3} = \frac{S_{\nu,2}(S_{\nu,4}C_{\nu,1}^2 - S_{\nu,1}C_{\nu,4}^2)}{\det(A)} \tag{43}$$

$$b_{\nu,4} = -\frac{S_{\nu,1}(S_{\nu,3}C_{\nu,2}^2 - S_{\nu,2}C_{\nu,3}^2)}{\det(A)}$$

where $\det(A)$ is again given by (40).

We then substitute these coefficients into (41) and define a regularization factor satisfying

$$\epsilon_g = \max\left\{\frac{L_g|b_{\nu,j}|}{a_{\nu,j}} \mid j \in \{1,2,3,4\}, \nu \in \{(0,1),(1,0)\}\right\}. \tag{44}$$

We will verify that this is finite and bounded in section 4.

3.4. Approximation of the nonlinear operator

We now have the building blocks in place to describe a discretization of the full nonlinear operator (8) at the point $x_i \in \mathcal{G}$.

The variational formulation of the modified determinant (30) requires performing a minimization over the set V of orthogonal bases for $\mathbb{R}^2$. In the discrete version, we consider a finite subset of V that ensures that all directions are resolved in the limit $h \to 0$. A simple choice is given by

$$\tilde{V} = \left\{(\cos\theta, \sin\theta) \mid \theta = jd\theta, \, j = 0, \ldots, \frac{\pi}{2d\theta}\right\} \tag{45}$$

where $d\theta = \mathcal{O}\left(\sqrt{h}\right)$ is the angular resolution of the stencil described in subsection 3.1.

The PDE involves several different functions of the gradient. For compactness, we introduce the shorthand notation

$$g_{1,\nu}(x_i, p) = \mathcal{D}_{\nu\nu}\tilde{c}(x_i, T(x_i, p)) \tag{46}$$

where the differencing is performed only in the first argument of $\tilde{c}$. This involves the explicit formulas for the optimal map given in (5)-(6).

We also define

$$g_2(x_i, p) = \frac{\left|\det D_{xy}^2 c\,(x_i, T(x_i, p))\right|}{f_2\,(T(x_i, p))} \tag{47}$$

recalling that the determinant of the mixed Hessian can be replaced with the simple explicit representations obtained in subsection 2.3, which we here denote by $H(p)$.

The discretization of these functions of the gradient require a regularization parameter (44), which involves Lipschitz bounds on these functions. We select bounds satisfying

$$L_{g_1} > L_{\tilde{c}} L_T \tag{48}$$

and

$$L_{g_2} \geq \left\|\frac{1}{f_2}\right\|^2 \left(\|f_2\| L_H + \|H\| L_{f_2} L_T\right) \tag{49}$$

where L_T is the Lipschitz constant of the optimal map $T(x, p)$ with respect to the variable p, L_H and L_{f_2} are the Lipschitz constants of the functions H and f_2 respectively, and

$$L_{\tilde{c}} = \max_{|v|=1, x, y \in \mathbb{S}^2} \left\| \nabla_y \left(v^T D_{xx}^2 \tilde{c}(x, y) v \right) \right\|. \tag{50}$$

We remark that while these constants L_{g_1} and L_{g_2} depend on the problem data and particular cost function, they are all guaranteed to be bounded under the assumptions of Hypothesis 2 and can be computed explicitly using the formulas in (5), (6), (27), (34), and (35).

We can then write down the full discretization of (8) as

$$\begin{aligned}
F^h(x, u(x) - u(\cdot)) = -\min_{(v_1, v_2) \in \tilde{V}} \prod_{j=1}^2 \max\left\{ \mathcal{D}_{v_j v_j} u(x_i) + \tilde{g}_{1,v_j}^-(\mathcal{D}u(x_i)), 0 \right\} \\
+ f_1(x_i)\tilde{g}_2^+\,(x_i, \mathcal{D}u(x_i)).
\end{aligned} \tag{51}$$

3.5. Solution method

In order to efficiently obtain a convergent approximation to (8), we will slightly modify the two-step procedure described in (52)-(25). We propose instead the following solution and verification process, which is equivalent.

1. Solve the discrete system

$$F^h(x, v^h(x) - v^h(\cdot)) + \sqrt{h}v^h(x) = 0, \quad x \in \mathcal{G} \tag{52}$$

  for the grid function v^h.
2. Verify that the grid function v^h satisfies the bounds

$$E^h(x, v^h(x) - v^h(\cdot)) \leq R, \quad x \in \mathcal{G}.$$

3. If the verification step fails, redefine v^h by solving the modified system (52)

$$G^h\left(x, v^h(x) - v^h(\cdot)\right) + \sqrt{h}v^h(x) = 0, \quad x \in \mathcal{G}$$

  using the solution obtained in Step 1 as an initial guess.
4. Define the discrete solution

$$u^h(x) = v^h(x) - v^h(x_h^*), \quad x \in \mathcal{G}. \tag{53}$$

We note that in practice, we have never found Step 3 above to be necessary. Thus although this procedure appears to be longer than simply performing Steps 3-4, it actually allows us to obtain the same solution by solving a simpler system.

The strong nonlinearity in the PDE (8), particularly in that it involves nonlinear gradient terms that have very little required structure, makes the construction of a nonlinear Gauss-Jacobi, algebraic multigrid, and/or approximate Newton-type method highly nontrivial. In the present work, we perform all our computations using explicit parabolic schemes of the form

$$v_{n+1}^h(x_i) = v_n^h(x_i) - \Delta t F^h(v_n^h(x_i), v_n^h(x_i) - v_n^h(\cdot)).$$

As discussed in [25], Δt has to satisfy a nonlinear CFL condition in order to guarantee convergence. In particular, we require $\Delta t < 1/L_{F^h}$, where L_{F^h} is the Lipschitz constant of F^h with respect to the arguments u_i^h. This Lipschitz constant scales like $L_{F^h} = \mathcal{O}(h^{-2})$ and can either be determined explicitly *a priori* or adaptively under a requirement that the residual should decrease. In some cases, acceleration of this process is possible using the approach of [29].

Faster solvers for these kinds of systems are, of course, desirable and will be explored in future work.

4. Convergence

We are now prepared to prove that the numerical method defined by (51) and the subsequent solution procedure converges.

Theorem 19 (Convergence). *Under the assumptions of Hypothesis 2, let $u \in C^3(\mathbb{S}^2)$ be the unique solution of (8) satisfying $u(x^*) = 0$. Let $\mathcal{G}^h$ be a grid satisfying Hypothesis 13 and let F^h be defined as in (51). Then for each sufficiently small $h > 0$, the grid function u^h defined in (53) is uniquely defined. Moreover, u^h converges uniformly to u as $h \to 0$.*

This result follows immediately from the framework described in Theorem 14 provided we can verify that our approximation scheme F^h is consistent (Lemma 26) and monotone (Lemma 28). This will be accomplished in several lemmas throughout the remainder of this section.

4.1. Bounds on coefficients

We begin by demonstrating that the coefficients $a_{v,j}, b_{v,j}$ appearing in the approximation of the second directional derivatives can be bounded.

Lemma 20 (Bounds on coefficients (second derivatives)). *There exists a constant $C > 0$ such that for all sufficiently small $h > 0$ and $v \in \mathbb{R}^2$, the coefficients defined by (39) satisfy*

$$a_{v,j} \geq \frac{C}{h}.$$

Proof. We establish a bound for the coefficient $a_{v,1}$; the remaining coefficients are similar.

Recall the notation $C_{v,j} = h_{v,j} \cos\theta_{v,j}$, $S_{v,j} = h_{v,j} \sin\theta_{v,j}$. Since each $x_{v,j}$ lies in the jth quadrant, each of these terms has a definite sign. Based on the requirements on $h_{v,j}, d\theta_{v,j}$ established in Lemma 18, we can record the asymptotic bounds

$$(r - 2h)\left(1 - \mathcal{O}(d\theta^2)\right) \leq |C_{v,j}| \leq r$$

$$(r - 2h)\left(d\theta - \mathcal{O}(d\theta^3)\right) \leq |S_{v,j}| \leq r\left(2d\theta + \mathcal{O}(d\theta^3)\right).$$

We recall also that $r, d\theta = \mathcal{O}(\sqrt{h})$.

These observations allow us to establish the following bounds:

$$
\begin{aligned}
\det(A) =& (C_{v,3}S_{v,2} - C_{v,2}S_{v,3})(C_{v,1}^2 S_{v,4} - C_{v,4}^2 S_{v,1}) \\
& - (C_{v,1}S_{v,4} - C_{v,4}S_{v,1})(C_{v,3}^2 S_{v,2} - C_{v,2}^2 S_{v,3}) \\
\leq& r^5(4d\theta + \mathcal{O}(d\theta^3))(4d\theta + \mathcal{O}(d\theta^3)) + r^5(4d\theta + \mathcal{O}(d\theta^3))(4d\theta + \mathcal{O}(d\theta^3)) \\
=& 32r^5 d\theta^2 + \mathcal{O}(h^{9/2})
\end{aligned}
$$

and

$$
\begin{aligned}
-2S_{v,4}(C_{v,2}S_{v,3} - C_{v,3}S_{v,2}) \geq & 2(r - 2h)^3(d\theta - \mathcal{O}(d\theta^3))\left(2d\theta - \mathcal{O}(d\theta^3)\right) \\
& = 4r^3 d\theta^2 + \mathcal{O}(h^3).
\end{aligned}
$$

Combining these bounds, we obtain

$$a_{\nu,1} \geq \frac{4r^3 d\theta^2 + \mathcal{O}(h^3)}{32r^5 d\theta^2 + \mathcal{O}(h^{9/2})} = \frac{1}{8r^2}\left(1 + \mathcal{O}(\sqrt{h})\right)$$

where $r^2 = \mathcal{O}(h)$. □

Lemma 21 (*Bounds on coefficients (first derivatives)*). *There exists a constant $C > 0$ such that for all sufficiently small $h > 0$ and $\nu \in \mathbb{R}^2$, the coefficients defined by* (42) *satisfy*

$$\left|b_{\nu,j}\right| \leq \frac{C}{\sqrt{h}}.$$

Proof. We proceed as in the proof of Lemma 20 and compute the bounds

$$\begin{aligned}
\det(A) =& (C_{\nu,3}S_{\nu,2} - C_{\nu,2}S_{\nu,3})(C_{\nu,1}^2 S_{\nu,4} - C_{\nu,4}^2 S_{\nu,1}) \\
& - (C_{\nu,1}S_{\nu,4} - C_{\nu,4}S_{\nu,1})(C_{\nu,3}^2 S_{\nu,2} - C_{\nu,2}^2 S_{\nu,3}) \\
\geq& \ 8(r-2h)^5 \left(1 - \mathcal{O}(d\theta^2)\right)^3 \left(d\theta - \mathcal{O}(d\theta^3)\right)^2 \\
=& \ 8r^5 d\theta^2 + \mathcal{O}(h^4)
\end{aligned}$$

and

$$0 \leq S_{\nu,4}(S_{\nu,3}C_{\nu,2}^2 - S_{\nu,2}C_{\nu,3}^2) \leq 2r^4(2d\theta + \mathcal{O}(d\theta^3))^2 = 8r^4 d\theta^2 + \mathcal{O}(h^4).$$

Combining these, we find that

$$0 \leq b_{\nu,1} \leq \frac{8r^4 d\theta^2 + \mathcal{O}(h^4)}{8r^5 d\theta^2 + \mathcal{O}(h^4)} = \frac{1}{r}\left(1 + \mathcal{O}(\sqrt{h})\right)$$

with $r = \mathcal{O}\left(\sqrt{h}\right)$.

The other coefficients are similar, though some are positive and some are negative. □

4.2. Bounds on Lipschitz constants

We next establish Lipschitz bounds on the functions $g_{1,\nu}, g_2$ defined in (46) and (47), which play an important role in the discretization of functions of the gradient.

Lemma 22 (*Lipschitz bound on $g_{1,\nu}$*). *Let $x_i \in \mathcal{G}$ be fixed. Then for $p \in \mathcal{T}_{x_i}$, the function*

$$g_{1,\nu}(p) = \mathcal{D}_{\nu\nu}\tilde{c}(x_i, T(x_i, p)),$$

has a Lipschitz constant L satisfying

$$L \leq L_{\tilde{c}}L_T + \mathcal{O}(\sqrt{h}).$$

Proof. We first recall that $g_{1,\nu}(p)$ involves a finite difference discretization. By consistency, we have

$$g_{1,\nu}(p) = \nu^T \left(D_{xx}^2 \tilde{c}(x_i, T(x_i, p))\right)\nu + C(p)\sqrt{h}$$

where the coefficient $C(p)$ arising in the discretization error is at least Lipschitz continuous in p. Then using regularity, we can calculate

$$\begin{aligned}
g_{1,\nu}(p) - g_{1,\nu}(q) &= \nu^T D_{xx}^2 \left(\tilde{c}(x_i, T(x_i, p)) - \tilde{c}(x_i, T(x_i, q))\right)\nu + (C(p) - C(q))\sqrt{h} \\
&\leq \max_{|\nu|=1, x, y \in \mathbb{S}^2} \left\|\nabla_y \left(\nu^T D_{xx}^2 \tilde{c}(x,y)\nu\right)\right\| \|T(x_i, p) - T(x_i, q)\| + \mathcal{O}(\sqrt{h})\|p - q\| \\
&\leq \left(L_{\tilde{c}}L_T + \mathcal{O}(\sqrt{h})\right)\|p - q\|. \quad \square
\end{aligned}$$

Lemma 23 *(Lipschitz bound on g_2). Let $x_i \in \mathcal{G}$ be fixed. Then for $p \in \mathcal{T}_{x_i}$, the function*

$$g_2(p) = \frac{\left|\det D_{xy}^2 \tilde{c}(x_i, T(x_i, p))\right|}{f_2(T(x_i, p))}$$

has a Lipschitz constant L satisfying

$$L \leq \left\| \frac{1}{f_2} \right\|^2 \left(\|f_2\| L_H + \|H\| L_{f_2} L_T \right).$$

Proof. Using the notation $H(p) = \left|\det D_{xy}^2 \tilde{c}(x_i, T(x_i, p))\right|$, we have

$$\begin{aligned}
g_2(p) - g_2(q) &= \frac{H(p)}{f_2(T(x_i, p))} - \frac{H(q)}{f_2(T(x_i, q))} \\
&= \frac{f_2(T(x_i, q))H(p) - f_2(T(x_i, p))H(q)}{f_2(T(x_i, p))f_2(T(x_i, q))} \\
&= \frac{f_2(T(x_i, q))(H(p) - H(q)) + H(q)(f_2(T(x_i, q)) - f_1(T(x_i, p)))}{f_2(T(x_i, p))f_2(T(x_i, q))} \\
&\leq \left\| \frac{1}{f_2} \right\|^2 \left(\|f_2\| L_H + \|H\| L_{f_2} L_T \right) \|p - q\|. \quad \square
\end{aligned}$$

4.3. Lax-Friedrichs approximations

We now verify that the Lax-Friedrichs type approximations for functions of the gradient defined in (41) are both consistent and monotone.

Lemma 24 *(Consistency of functions of gradient). Let g be Lipschitz continuous with Lipschitz constant L_g and $\phi \in C^2$. Then*

$$\begin{aligned}
\tilde{g}^{\pm}(\mathcal{D}\phi(x_i)) = {}& g\left(\sum_{v \in \{(1,0),(0,1)\}} v \sum_{j=1}^4 b_{v,j} \left(\phi(x_{v,j}) - \phi(x_i) \right) \right) \\
& \mp \epsilon_g \sum_{v \in \{(1,0),(0,1)\}} \sum_{j=1}^4 a_{v,j} \left(\phi(x_{v,j}) - \phi(x_i) \right),
\end{aligned}$$

is a consistent approximation of $g(\nabla\phi)$.

Proof. We note that by construction, we have

$$\lim_{h \to 0} \sum_{v \in \{(1,0),(0,1)\}} v \sum_{j=1}^4 b_{v,j} \left(\phi(x_{v,j}) - \phi(x_i) \right) = \nabla\phi(x_i)$$

and

$$\lim_{h \to 0} \sum_{v \in \{(1,0),(0,1)\}} \sum_{j=1}^4 a_{v,j} \left(\phi(x_{v,j}) - \phi(x_i) \right) = \Delta\phi(x_i).$$

Using Lemmas 20-21, we can also bound ϵ_g via

$$\begin{aligned}
\epsilon_g &= \max \left\{ \frac{L_g |b_{v,j}|}{a_{v,j}} \mid j \in \{1, 2, 3, 4\}, \, v \in \{(0, 1), (1, 0)\} \right\} \\
&\leq \frac{L_g (C_b / \sqrt{h})}{C_a / h} \\
&= C \sqrt{h}
\end{aligned}$$

where the constant C is independent of h.

Combining these results, we obtain

$$\lim_{h \to 0} \tilde{g}^{\pm}(\mathcal{D}\phi(x_i)) = g(\nabla\phi(x_i)),$$

with a truncation error of $\mathcal{O}(\sqrt{h})$. □

Lemma 25 (*Monotonicity of functions of the gradient*). *Let g be Lipschitz continuous with Lipschitz constant L_g and $\phi \in C^2$. Then the schemes $\tilde{g}^+(\mathcal{D}\phi(x_i))$ and $\tilde{g}^-(\mathcal{D}\phi(x_i))$ are monotone and negative monotone respectively.*

Proof. We verify monotonicity of $\tilde{g}^+(\mathcal{D}\phi(x_i))$; the other part of the argument is identical.

Denoting by d_j the differences $\phi(x_i) - \phi(x_{\nu,j})$ allows us to express the scheme more compactly as

$$G(d) = g\left(- \sum_{\nu \in \{(1,0),(0,1)\}} \nu \sum_{j=1}^{4} b_{\nu,j} d_j \right) + \epsilon_g \sum_{\nu \in \{(1,0),(0,1)\}} \sum_{j=1}^{4} a_{\nu,j} d_j.$$

We now introduce a perturbation $\delta > 0$ into the argument d_k to obtain

$$G(d + \delta \hat{d}_k) - G(d) \geq -L_g \left| b_{\nu,k} \right| \delta + \epsilon_g a_{\nu,k} \delta$$
$$\geq 0$$

since $\epsilon_g \geq L \left| b_{\nu,k} \right| / a_{\nu,k}$. Therefore the scheme is monotone. □

4.4. Consistency and monotonicity

We now establish consistency and monotonicity of the approximation (51), which in turn establishes the convergence result (Theorem 19).

Lemma 26 (*Consistency*). *Under the assumptions of Hypotheses 2, 13, the scheme F^h defined by (51) is consistent with the PDE (8) on the space of C^2 functions satisfying the constraint (9).*

Proof. Let $\phi \in C^2$ satisfy the constraint (9). Then the determinant in (8) can be equivalently expressed as in (31):

$$\det{}^+(D^2\phi(x) + A(x, \nabla\phi(x))) = \min_{\nu_1, \nu_2 \in V} \prod_{j=1}^{2} \max\left\{ \frac{\partial^2 \phi(x)}{\partial \nu_j^2} + \left. \frac{\partial^2 c(x,y)}{\partial \nu_j^2} \right|_{y = T(x, \nabla\phi(x))}, 0 \right\}.$$

By design and by Lemma 24, the components of the PDE are approximated consistently (with truncation error $\mathcal{O}(\sqrt{h})$). That is,

$$\lim_{h \to 0} \left\{ \mathcal{D}_{\nu_j \nu_j}\phi(x_i) + \tilde{g}_{1,\nu_j}^-(\mathcal{D}\phi(x_i)) \right\} = \frac{\partial^2 \phi(x)}{\partial \nu_j^2} + \left. \frac{\partial^2 c(x,y)}{\partial \nu_j^2} \right|_{y = T(x, \nabla\phi(x))}$$

and

$$\lim_{h \to 0} \tilde{g}_2^+ (x_i, \mathcal{D}\phi(x_i)) = \frac{\left| \det D_{xy}^2 c(x_i, T(x_i, \nabla\phi(x_i))) \right|}{f_2(T(x_i, \nabla\phi(x_i)))}.$$

We recall that the maximum and minimum operators are continuous, $f_1 \in C^1$, and $\tilde{V}$ is a consistent approximation of the set V with angular resolution $\mathcal{O}(d\theta) = \mathcal{O}(\sqrt{h})$. Thus the combinations of these operators in the scheme F^h satisfy

$$\lim_{h \to 0, x_i \to x} - \min_{(\nu_1, \nu_2) \in \tilde{V}} \prod_{j=1}^{2} \max\left\{ \mathcal{D}_{\nu_j \nu_j}\phi(x_i) + \tilde{g}_{1,\nu_j}^-(\mathcal{D}\phi(x_i)), 0 \right\} + f_1(x_i)\tilde{g}_2^+ (x_i, \mathcal{D}\phi(x_i))$$

$$= -\det{}^+(D^2\phi(x) + A(x, \nabla\phi(x))) + \left| \det D_{xy}^2 c(x, T(x, \nabla\phi(x))) \right| \frac{f_1(x)}{f_2(T(x, \nabla\phi(x)))},$$

which establishes consistency. □

Corollary 27 (*Truncation error*). *Under the assumptions of Hypotheses 2, 13, the scheme F^h defined by (51) has local truncation error $\tau(h) = \mathcal{O}(\sqrt{h})$.*

Lemma 28 *(Monotonicity). Under the assumptions of Hypotheses 2, 13, the scheme F^h defined by (51) is monotone.*

Proof. By construction (and see Lemma 25), the schemes for $\mathcal{D}_{\nu\nu}\phi$ and $\tilde{g}_{1,\nu}^-(\mathcal{D}u)$ are negative monotone. Addition and the maximum function preserve this so that

$$\max\left\{\mathcal{D}_{\nu\nu}u(x_i) + \tilde{g}_{1,\nu}^-(\mathcal{D}u(x_i)), 0\right\}$$

is also negative monotone for any $\nu \in \mathbb{R}^2$. Since this is also non-negative, products of these terms preserve the negative monotonicity so that

$$\min_{(\nu_1,\nu_2)\in\tilde{V}} \prod_{j=1}^{2} \max\left\{\mathcal{D}_{\nu_j\nu_j}u(x_i) + \tilde{g}_{1,\nu_j}^-(\mathcal{D}u(x_i)), 0\right\}$$

is also negative monotone.

We recall also that f_1 is non-negative and $\tilde{g}_2^+$ is monotone (Lemma 25). Therefore the full scheme

$$-\min_{(\nu_1,\nu_2)\in\tilde{V}} \prod_{j=1}^{2} \max\left\{\mathcal{D}_{\nu_j\nu_j}u(x_i) + \tilde{g}_{1,\nu_j}^-(\mathcal{D}u(x_i)), 0\right\} + f_1(x_i)\tilde{g}_2^+(x_i, \mathcal{D}u(x_i))$$

is monotone. $\square$

4.5. Extensions to nonsmooth problems

The results of [22] ensure existence of weak solutions to the optimal transport PDE (8) in the relaxed setting where $f_1, f_2 \in L^1$ with f_2 bounded away from zero and f_1 bounded away from infinity. In [20], the authors of this article showed that these solutions can be computed for the squared geodesic cost as in Theorem 14 if the scheme F^h is additionally required to underestimate when applied to the true solution. This result can now been extended to the regularized logarithmic cost via Remark 17.

Naturally underestimating schemes can be constructed for Monge-Ampère type equations in some cases [7,10]. An alternate approach is to utilize a scheme of the form

$$\tilde{F}^h(x, u(x) - u(\cdot)) = F^h(x, u(x) - u(\cdot)) - h^\alpha$$

for a sufficiently small $\alpha > 0$. This preserves the consistency and monotonicity of the original scheme, while decreasing the value of the scheme and forcing it to be negative when applied to the true solution of the PDE.

In addition, definition of consistency (Definition 6) in terms of upper and lower semicontinuous envelopes of the PDE operator allows us to accommodate discontinuous data f_1, f_2 as in [10]. This does not require any change in the way $f_1(x)$ is handled. However, functions $f_2 \notin C^{0,1}$ must be carefully regularized to preserve consistency and monotonicity since it takes as its argument terms $T(x, \nabla u(x))$ that involve gradients.

The approach we propose is to introduce a discrete version of the target density function

$$f_2^h(y) = (K_{h^{1/4}} * f_2)(y) \tag{54}$$

where $K_{h^{1/4}}$ is a mollifier that ensures that the Lipschitz constant of f_2^h satisfies

$$L_{f_2^h} \leq h^{-1/4}.$$

From here, we use the same discretization of $f_2(T(x, \nabla u(x)))$ introduced in (41). We note that in this case (following Lemma 24), the regularization parameter will satisfy

$$\epsilon_{g_2^h} = \mathcal{O}\left(L_{f_2^h}\sqrt{h}\right) = \mathcal{O}(h^{1/4}).$$

Since this parameter converges to zero as $h \to 0$, the resulting scheme will still be consistent in the sense of Definition 6. The monotonicity result (Lemma 25) is unchanged.

5. Computational results

5.1. Structured and unstructured grids

The scheme we have built works well on both structured and unstructured grids, provided that they satisfy the mild conditions of Hypothesis 13. By structured we mean that there exists a deterministic way of building the grid. Likewise, unstructured here means there is a stochastic element in the construction of the grid.
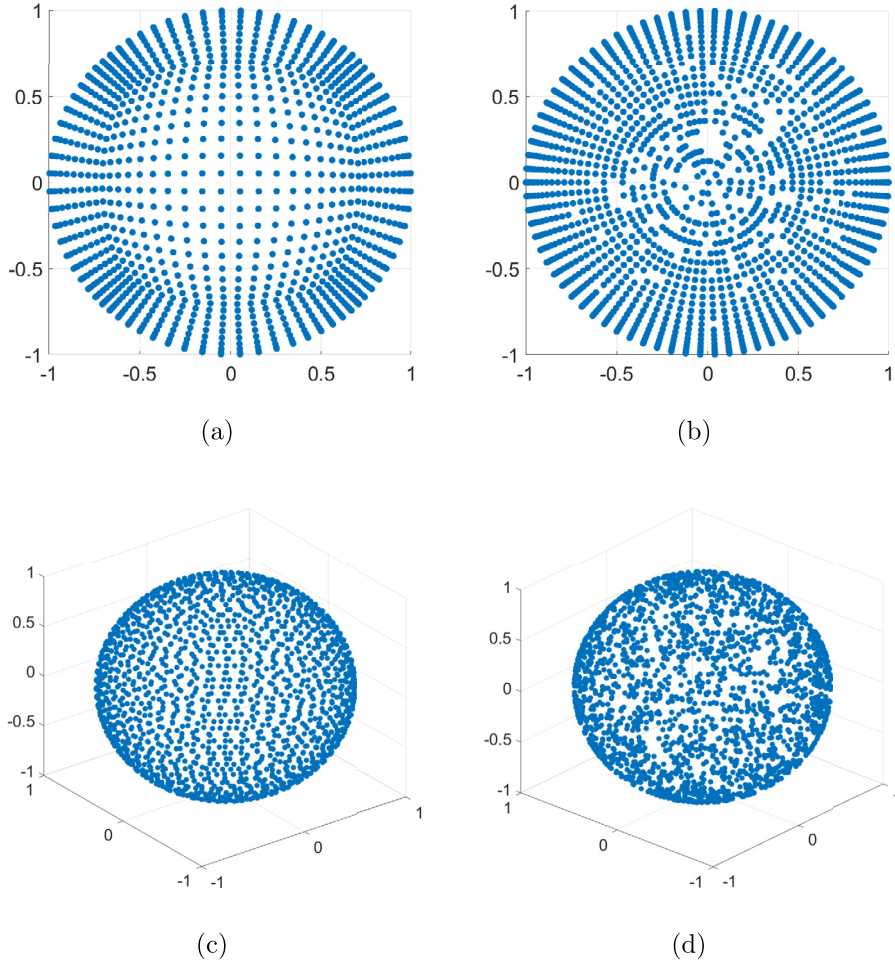
Fig. 5. Top down views of a (a) cube grid and (b) latitude-longitude grid. Side views of a (c) layered grid and a (d) randomgrid.

Here, we describe four types of grids that satisfy these hypotheses and that we make use of in practice: the cube grid (structured), the random grid (fully unstructured), the latitude-longitude grid (unstructured), and the layered grid (structured).

The structured cube grid is constructed as follows. First, a grid of evenly-spaced points is generated on the faces of a cube which contains the sphere. Then, the points on such a grid on the cube are projected onto the sphere. See Fig. 5(a) for an example of the resulting grid.

The semi-unstructured latitude-longitude grid is constructed as follows. We begin with a structured grid composed of points equally spaced in both latitude θ and longitude ϕ. However, this produces a highly over-resolved grid near the poles, which does not satisfy the required structure conditions. In order to get rid of this redundancy, we stochastically remove grid points within a geodesic distance $\mathcal{O}(h)$ of both poles. That is, for each grid point $x_i = (\theta_i, \phi_i)$, we compute $\xi_i = \sin\theta_i$ and generate a value using the random variable $\Xi \sim \text{Unif}([0, 1])$. If $\Xi > \xi_i$, then we remove the point (θ_i, ϕ_i) from the grid. The removal of these points creates a new unstructured grid that almost surely satisfies Hypothesis 13. See Fig. 5(b) for an example of such a grid.

To construct the structured layered grid, we take an integer n and $\epsilon = \mathcal{O}(h)$ and define the rows: $\text{row}_j = \epsilon + j\dfrac{\pi - 2\epsilon}{n}$.

For each row, we have $\text{col}_{kj} = j\phi_0 + k\dfrac{2\pi}{\lfloor n * \sin\theta_n \rfloor}$, where $\phi_0 = \dfrac{1 + \sqrt{5}}{2}$, which is the golden ratio. This creates a nice spread of points inspired by the seed packing of sunflowers. Then, we introduce the grid points $x_{jk} = (\theta, \phi) = (\text{row}_j, \text{col}_{jk})$ for $j = 1, \ldots, n$ and $k = 1, \ldots, \text{floor}\{n * \sin\theta_n\}$. This grid, by construction, will satisfy Hypothesis 13. See Fig. 5(c).

Finally, we consider a fully unstructured random grid. After defining the set

$$R = \{(x, y) \in [0, \pi] \times [0, 1] : y \leq \sin(x)\}$$
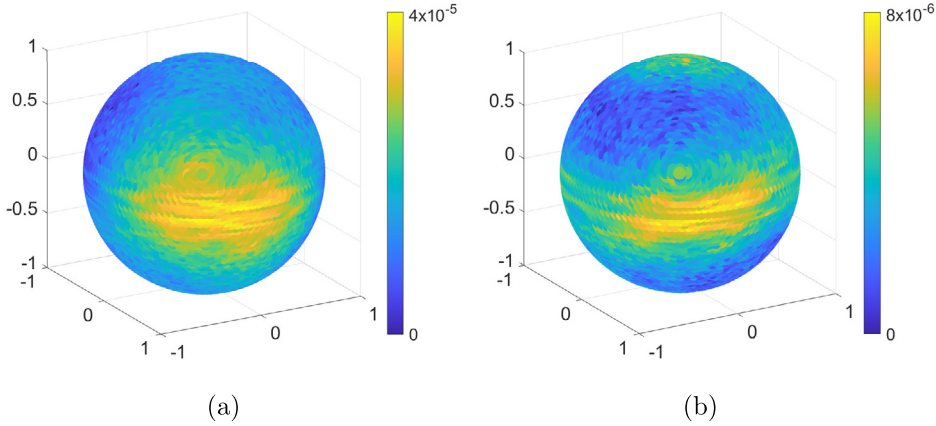
and the projection

Fig. 6. Solutions obtained for the (a) squared geodesic cost and (b) logarithmic cost when $f_1 = f_2$ on a layered grid consisting of $N = 7722$ points.

$$P(x, y) = x,$$

we sample $\Phi \sim \text{Unif}([0, 2\pi])$ and $\tilde{\Theta} \sim \text{Unif}(R)$, then define $\Theta = P(\tilde{\Theta})$.

We take as grid points the random variables

$$(X, Y, Z) = (\sin \Theta \cos \Phi, \sin \Theta \sin \Phi, \cos \Theta),$$

which satisfy the conditions that $(X, Y, Z) \sim \text{Unif}(\mathbb{S}^2)$. The resulting grid almost surely satisfies Hypothesis 13. See Fig. 5(d) for an example of such a grid.

5.2. Recovering constant solutions

For both cost functions, in the case where $f_1 = f_2$, the resulting solution will be a constant function ($u(x) = 0$). Fig. 6 shows the solutions obtained for both the squared geodesic and logarithmic cost functions. Importantly, these are effectively constant to within a tolerance less than the expected consistency error ($\tau(h) = \mathcal{O}\left(N^{-1/4}\right) \approx 0.1$) of the method.

5.3. Small perturbation

Here we demonstrate with computations for the squared geodesic cost what happens when the target mass density f_2 is obtained through a slight perturbation (a rotation through an angle θ) of the source mass density f_1. In particular, we choose the density functions

$$\begin{cases} f_1(x, y, z) = \frac{1}{4\pi - 4}\left(1 - 0.5 \cos\left(\frac{\pi}{2}x\right)\right) \\ f_2(x, y, z) = \frac{1}{4\pi - 4}\left(1 - 0.5 \cos\left(\frac{\pi}{2}(x \cos\theta + y \sin\theta)\right)\right) \end{cases} \tag{55}$$

This problem has the flavor of a translation. In the Euclidean setting, translations are exact solutions of the optimal transport problem. However, it is not the case that rotations are exact solutions on the sphere. See Fig. 7 for a top-view of the computed gradient map. In particular, we observe that the bulk of the mass does undergo a clockwise rotation. However, in order to conserve mass and regularity, there is also backwards "flow" observed in areas of low density (top and bottom of the figure).

5.4. Comparing structured and unstructured grids

To show the robustness of our generalized finite difference scheme with respect to the structure of the grid, we next show a side-by-side comparison of solutions obtained using a fully structured layered grid and a fully unstructured random grid. We choose a non-smooth source density f_1 (which concentrates mass at the poles) and constant target density f_2:

$$\begin{cases} f_1(\theta, \phi) = (1 - \epsilon)\frac{1}{5.8735}\left(\theta - \frac{\pi}{2}\right)^2 + \frac{\epsilon}{4\pi} \\ f_2(\theta, \phi) = \frac{1}{4\pi} \end{cases} \tag{56}$$

for $\epsilon = 0.1$.

See Fig. 8 for the computed solutions, which are identical to within a tolerance well under the expected consistency error ($\tau(h) \approx 0.1$). Moreover, in both solutions we observe u attaining minima at the poles, with a gradient directed away from the poles. This has the expected result of transporting mass out from the poles, where it was initially concentrated in the source f_1.
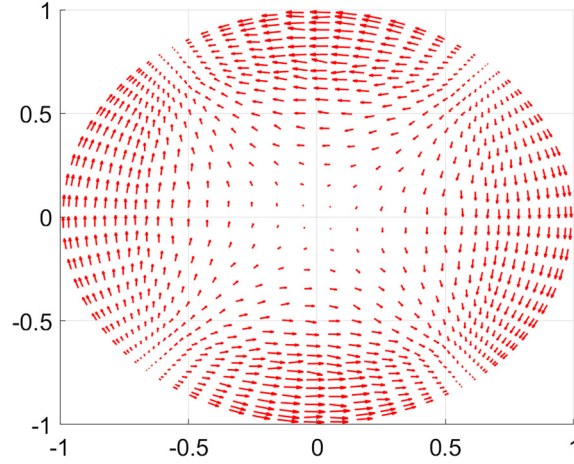
Fig. 7. Top view of the local gradient obtained numerically when f_2 is obtained from f_1 through a small rotation. The solution was computed using a cube grid with $N = 2168$ points.
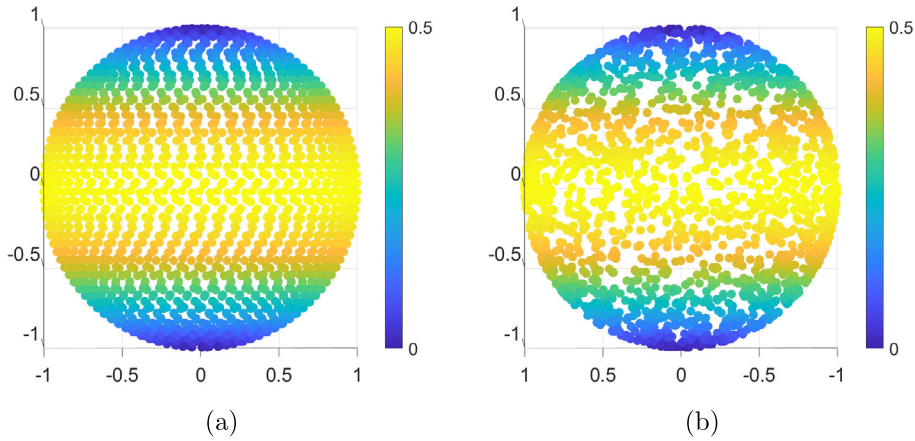

Fig. 8. Solution u obtained for densities (56) on $N = 2006$ point (a) layered and (b) random grids.

5.5. Mesh generation

In the context of mesh generation, one of the most important aspects of the squared geodesic cost is the lack of tangling in mesh restructuring. To demonstrate that our method preserves this critical property, we construct an explicit mesh and preserve the edge connections under the computed transport map T. We begin with a structured cube grid, where edges are obtained from the edges of the original cube grid that was projected onto the sphere. We select density functions with the goal of producing a transport map T that will concentrate mesh points around the equator.

$$\begin{cases} f_1(\theta, \phi) = \frac{1-\epsilon}{5.8735} \left(\theta - \frac{\pi}{2}\right)^2 + \frac{\epsilon}{4\pi} \\ f_2(\theta, \phi) = \frac{1}{4\pi} \end{cases} \tag{57}$$

where $\epsilon = 0.1$.

The resulting mesh restructuring is pictured in Fig. 9. As desired, the final mesh concentrates grid points around the equator without introducing any tangling into the mesh.

5.6. Logarithmic cost with full support

For the reflector antenna problem, previous available methods have successfully performed computations when the mass densities f_1 and f_2 have support on a subset of $\mathbb{S}^2$; see [19]. This does not allow the reflector antenna problem to be solved for cases where the source light intensity is, for example, omnidirectional. Here we demonstrate the computation for the case where both the mass densities have full support.
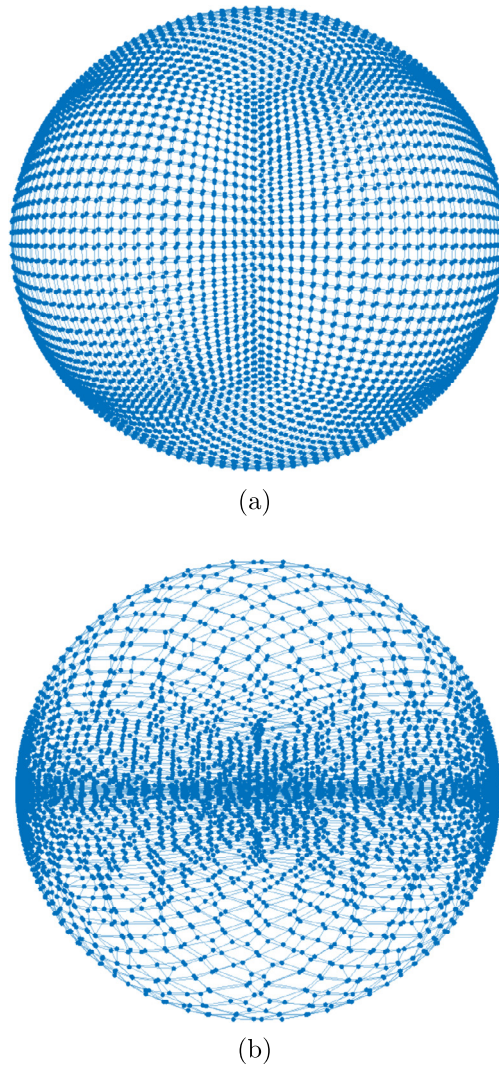

(a)

(b)

Fig. 9. (a) The original $N = 5048$ point cube mesh and (b) the mesh obtained by transporting through the computed optimal transport map.

Since applications in optics can often involve highly non-smooth densities, we take as our source density the image of a world map, which is mapped to a constant target density. See Fig. 10(a). Despite the potential singularity in the cost function and the non-smoothness of the problem data and solution, our method handles this problem with ease.

A simple illustration demonstrates the fact that the computed mapping T behaves as we would expect for the reflector antenna problem. Light emitted from the origin in the direction $x \in \mathbb{S}^2$ and reflecting off a very flat reflector would be transmitted to the opposite direction. Thus, we expect the optimal mapping to be an approximate inversion of the original globe. This intuition is supported by the nature of the logarithmic cost $c(x, y) = -\log \|x - y\|$, which has a singularity when $x = y$ and will thus tend to drive the target mass away from the source mass.

To illustrate this behavior, we assign different colors to different parts of the globe (oceans, continents in northern hemisphere, and continents in southern hemisphere), then plot the image of these points under the computed map T. See Fig. 10. As expected, we observe an approximate inversion of the continents. In addition, the oceans here have been chosen to have higher density than the continents in the source density. This means that the optimal transport mapping onto a uniform density should display a "flooding" of the continents, which is what we observe in our computed map.

5.7. Nonsmooth examples

Finally, we present the results of a computation (using the squared geodesic cost) where f_1 is unbounded and f_2 is not Lipschitz. Recall that this is a situation where the solution u is only guaranteed to be $C^1(\mathbb{S}^2)$ and the non-Lipschitz property of f_2 can easily cause issues regarding monotonicity and consistency. However, these issues can be resolved using the ideas in subsection 4.5. The density functions are given by
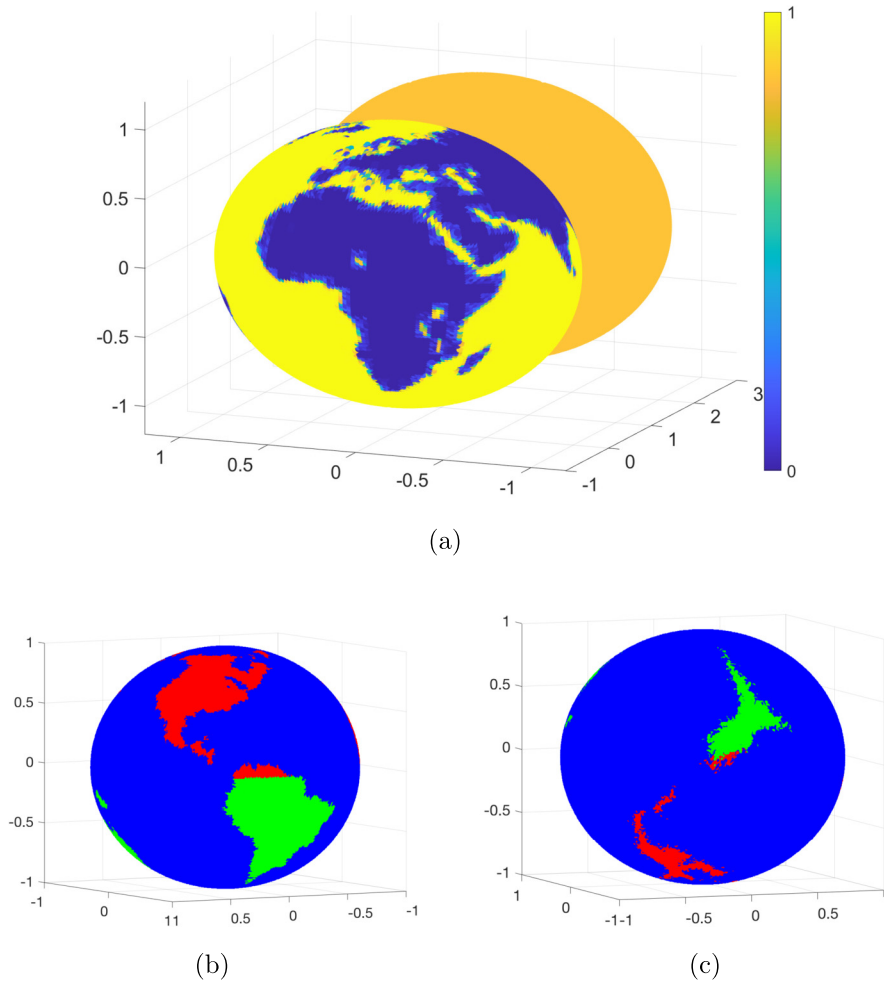
(a)

(b)

(c)

Fig. 10. (a) Density functions for the world map given on a $N \approx 20,000$ point latitude-longitude grid. (b) Domain with oceans labeled in blue, continental points in the northern hemisphere labeled in red, and continental points in the southern hemisphere labeled in green. (c) Image of the domain under the computed map T using the logarithmic cost.

$$\begin{cases} f_1(\theta, \phi) = \frac{1}{2\pi \cdot 1.86691} \theta^{-1/4} \\ f_2(\theta, \phi) = (1 - \epsilon) \frac{\theta^{3/4}}{17.2747} + \frac{\epsilon}{4\pi}, \end{cases} \tag{58}$$

where $\epsilon = 0.5$.

Despite the very strong singularities present in this example, our numerical method has no difficulty computing a solution. The density functions and computed gradient are shown in Fig. 11. As expected, we observe mass being transported downward away from the singularity.

6. Conclusion

We have introduced a provably convergent numerical method for solving the Optimal Transport problem on the sphere. The method applies to both the squared geodesic cost function (important in mesh generation) and the logarithmic cost function (arising in the reflector antenna problem). This method involves replacing the PDE on the manifold with equivalent PDEs on local tangent planes using a careful choice of geodesic normal coordinates. The resulting PDE was approximated using monotone schemes inspired by recent methods for solving the Monge-Ampère equation. However, significant new techniques were introduced to handle the complicated nonlinear combination of gradient and Hessian terms. Moreover, we introduced a smooth cutoff for the logarithmic cost function to enable well-posed, consistent, and monotone discretizations even in the presence of this potentially unbounded cost function.

We produced an implementation of this method and demonstrated via computations how our numerical method is able to address issues heretofore unresolved. Notably, the method can handle both structured and unstructured grids, non-smooth data, and produces compelling results in examples closely related to both the moving mesh and reflector antenna design problems.
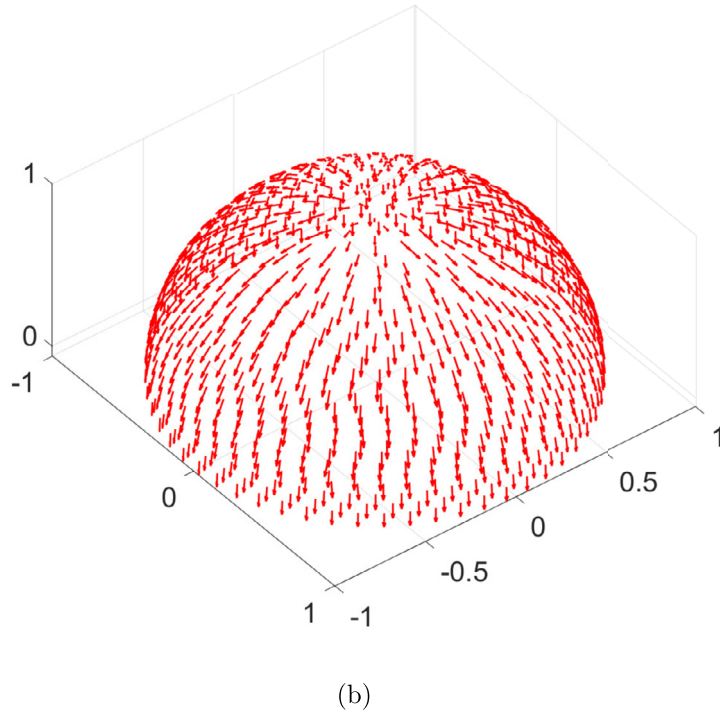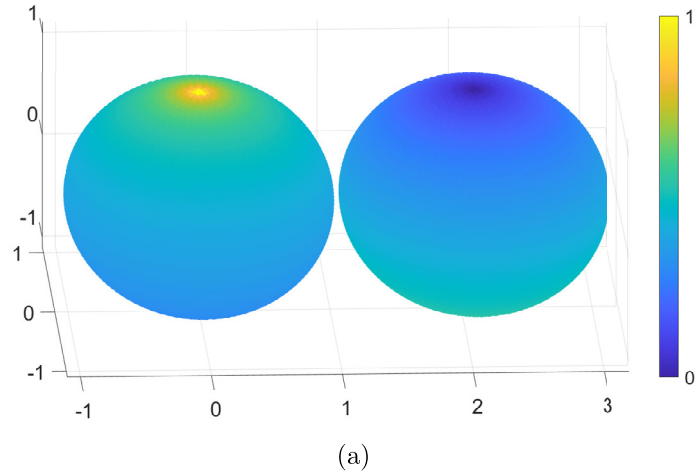
(a)

(b)

Fig. 11. (a) Unbounded source and non-Lipschitz target densities and (b) resulting gradient (visualized for the northern hemisphere) computed on a $N = 7722$ point layered grid.

CRediT authorship contribution statement

Brittany Froese Hamfeldt: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. **Axel G.R. Turnquist:** Funding acquisition, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Derivation of the mixed Hessian

In this appendix, we fill in the details of the derivation of simple expressions for the determinant of the mixed Hessian. For each cost function, we take the following approach:

1. Introduce orthogonal perturbations $\Delta p_1, \Delta p_2 \in \mathcal{T}_x$ such that $\Delta p_1 \cdot p = 0$ and $\Delta p_2 = \hat{p} \| \Delta p_2 \|$.
2. Establish that $T(x, p) - T(x, p + \Delta p_1)$ and $T(x, p) - T(x, p + \Delta p_2)$ are orthogonal to leading order.
3. Compute the change of area formula

$$
\begin{aligned}
&\left| \det \left(D_p T(x, p) \right) \right| \\
&= \lim_{\|\Delta p_1\|, \|\Delta p_2\| \to 0} \frac{d_{\mathbb{S}^2} \left(T(x, p), T(x, p + \Delta p_1) \right) d_{\mathbb{S}^2} \left(T(x, p), T(x, p + \Delta p_2) \right)}{\|\Delta p_1\| \|\Delta p_2\|} \\
&= \lim_{\|\Delta p_1\|, \|\Delta p_2\| \to 0} \frac{\| T(x, p) - T(x, p + \Delta p_1) \| \| T(x, p) - T(x, p + \Delta p_1) \|}{\|\Delta p_1\| \|\Delta p_2\|}
\end{aligned}
$$

where we can simplify the formulas by using the fact that

$$
\| T(x, p) - T(x, p + \Delta p) \| = d_{\mathbb{S}^2} \left(T(x, p), T(x, p + \Delta p) \right) + \mathcal{O} \left(\| \Delta p \|^2 \right).
$$

A.1. Squared geodesic cost

We begin with the squared geodesic cost, recalling that the mapping T has the explicit form

$$
T(x, p) = \cos \left(\| p \| \right) x + \sin \left(\| p \| \right) \frac{p}{\| p \|}.
$$

First consider a perturbation satisfying $\Delta p_1 \cdot p = 0$.

$$
\begin{aligned}
T(x, p) - T(x, p + \Delta p_1) =\ & x \left(\cos \| p \| - \cos \| p + \Delta p_1 \| \right) \\
& + \frac{p}{\| p \|} \sin \| p \| - \frac{p + \Delta p_1}{\| p + \Delta p_1 \|} \sin \| p + \Delta p_1 \|
\end{aligned}
$$

Now, since p and Δp_1 are orthogonal,

$$
\| p + \Delta p_1 \| = \sqrt{\| p \|^2 + \| \Delta p_1 \|^2} = \| p \| + \mathcal{O} \left(\| \Delta p_1 \|^2 \right).
$$

Thus to leading order we obtain

$$
T(x, p) - T(x, p + \Delta p_1) = \Delta p_1 \frac{\sin \| p \|}{\| p \|} + \mathcal{O} \left(\| \Delta p_1 \|^2 \right).
$$

Now, suppose that $\Delta p_2 = \| \Delta p_2 \| \frac{p}{\| p \|}$. In this case, $\| p + \Delta p_2 \| = \| p \| + \| \Delta p_2 \|$. As before, we compute to leading order:

$$
\begin{aligned}
& T(x, p) - T(x, p + \Delta p_2) \\
&= x \left(\cos \| p \| - \cos \| p + \Delta p_2 \| \right) + \frac{p}{\| p \|} \sin \| p \| - \frac{p + \Delta p_2}{\| p + \Delta p_2 \|} \sin \| p + \Delta p_2 \| \\
&= x \| \Delta p_2 \| \sin \| p \| - \frac{p}{\| p \|} \| \Delta p_2 \| \cos \| p \| + \frac{p}{\| p \|^2} \| \Delta p_2 \| \sin \| p \| - \frac{\Delta p_2}{\| p \|} \sin \| p \| \\
&\quad + \mathcal{O} \left(\| \Delta p_2 \|^2 \right) \\
&= x \| \Delta p_2 \| \sin \| p \| - \frac{p}{\| p \|} \| \Delta p_2 \| \cos \| p \| + \mathcal{O} \left(\| \Delta p_2 \|^2 \right).
\end{aligned}
$$

Now since $x \cdot \Delta p_1 = p \cdot \Delta p_1 = \Delta p_1 \cdot \Delta p_2 = 0$, we can easily verify that

$$
\left(T(x, p) - T(x, p + \Delta p_1) \right) \cdot \left(T(x, p) - T(x, p + \Delta p_2) \right) = o \left(\| \Delta p_1 \| + \| \Delta p_2 \| \right)
$$

so that the perturbations in the map are indeed orthogonal to leading order.

Now we can use orthogonality to easily compute the magnitudes of these perturbations via

$$
\| T(x, p) - T(x, p + \Delta p_1) \|^2 = \| \Delta p_1 \|^2 \frac{\sin^2 \| p \|}{\| p \|^2} + o \left(\| \Delta p_1 \|^2 \right)
$$

and

$$\|T(x, p) - T(x, p + \Delta p_2)\|^2 = \|\Delta p_2\|^2 \sin^2 \|p\| + \|\Delta p_2\|^2 \cos^2 \|p\| + o\left(\|\Delta p_2\|^2\right)$$

$$= \|\Delta p_2\|^2 + o\left(\|\Delta p_2\|^2\right),$$

where we have used the fact that x and $p/\|p\|$ are unit vectors.

Then we can compute the change of area formula as

$$|\det(D_p T(x, p))|$$

$$= \lim_{\|\Delta p_1\|, \|\Delta p_2\| \to 0} \frac{\|T(x, p) - T(x, p + \Delta p_1)\| \|T(x, p) - T(x, p + \Delta p_2)\|}{\|\Delta p_1\| \|\Delta p_2\|}$$

$$= \lim_{\|\Delta p_1\|, \|\Delta p_2\| \to 0} \frac{\|\Delta p_1\| \|\Delta p_2\| \sin\|p\|/\|p\| + o(\|\Delta p_1\| \|\Delta p_2\|)}{\|\Delta p_1\| \|\Delta p_2\|}$$

$$= \frac{\sin\|p\|}{\|p\|}.$$

Hence, the determinant of the mixed Hessian for the squared geodesic cost is

$$\left|\det D_{xy}^2 c(x, y)\right| = \frac{\|p\|}{\sin\|p\|}. \tag{A.1}$$

A.2. Logarithmic cost

Now we perform the same procedure for the logarithmic map, which has the explicit form

$$T(x, p) = x\frac{\|p\|^2 - 1/4}{\|p\|^2 + 1/4} - \frac{p}{\|p\|^2 + 1/4}.$$

We again begin with a perturbation satisfying $p \cdot \Delta p_1 = 0$ so that

$$\|p + \Delta p_1\| = \|p\| + \mathcal{O}\left(\|\Delta p_1\|^2\right).$$

To leading order, we can compute

$$T(x, p) - T(x, p + \Delta p_1) = x\frac{\|p\|^2 - 1/4}{\|p\|^2 + 1/4} - x\frac{\|p + \Delta p_1\|^2 - 1/4}{\|p + \Delta p_1\|^2 + 1/4}$$

$$- \frac{p}{\|p\|^2 + 1/4} + \frac{p + \Delta p_1}{\|p + \Delta p_1\|^2 + 1/4}$$

$$= \frac{\Delta p_1}{\|p^2\| + 1/4} + \mathcal{O}\left(\|\Delta p_1\|^2\right).$$

Next we consider an orthogonal perturbation $\Delta p_2 = \|\Delta p_2\| \frac{p}{\|p\|}$ so that

$$\|p + \Delta p_2\|^2 = \|p\|^2 + 2\|p\|\|\Delta p_2\| + \mathcal{O}\left(\|\Delta p_2\|^2\right).$$

Now we can compute to leading order

$$\|T(x, p) - T(x, p + \Delta p_2)\| = x\frac{\|p\|^2 - 1/4}{\|p\|^2 + 1/4} - x\frac{\|p\|^2 + 2\|p\|\|\Delta p_2\| - 1/4}{\|p\| + 2\|p\|\|\Delta p_2\|^2 + 1/4}$$

$$- \frac{p}{\|p\|^2 + 1/4} + \frac{p + \Delta p_2}{\|p\|^2 + 2\|p\|\|\Delta p_2\| + 1/4} + \mathcal{O}\left(\|\Delta p_2\|^2\right)$$

$$= -x\frac{2\|p\|\|\Delta p_2\|}{\|p\|^2 + 1/4} + x\frac{2\|p\|\|\Delta p_2\|(\|p\|^2 - 1/4)}{(\|p\|^2 + 1/4)^2}$$

$$+ \frac{\Delta p_2}{\|p\|^2 + 1/4} - p\frac{2\|p\|\|\Delta p_2\|}{(\|p\|^2 + 1/4)^2} + \mathcal{O}\left(\|\Delta p_2\|^2\right)$$

$$= -x\frac{\|p\|\|\Delta p_2\|}{(\|p\|^2 + 1/4)^2} + \hat{p}\frac{\|\Delta p_2\|(1/4 - \|p\|^2)}{(\|p\|^2 + 1/4)^2} + \mathcal{O}\left(\|\Delta p_2\|^2\right).$$

Since x, p, and $\triangle p_1$ are mutually orthogonal, we can immediately verify that

$$(T(x, p) - T(x, p + \triangle p_1)) \cdot (T(x, p) - T(x, p + \triangle p_2)) = o\left(\|\triangle p_1\| + \|\triangle p_2\|\right)$$

so that the perturbations in the mapping are again orthogonal to leading order.

Next we compute the lengths of the perturbations using orthogonality:

$$\|T(x, p) - T(x, p + \triangle p_1)\| = \|\triangle p_1\| \frac{1}{\|p\|^2 + 1/4} + o(\|\triangle p_1\|)$$

and

$$\begin{aligned} \|T(x, p) - T(x, p + \triangle p_2)\|^2 &= \frac{\|p\|^2\|\triangle p_2\|^2 + \|\triangle p_2\|^2(1/4 - \|p\|^2)^2}{(\|p\|^2 + 1/4)^4} + o(\|\triangle p_2\|^2) \\ &= \frac{\|\triangle p_2\|^2}{(\|p\|^2 + 1/4)^4}\left(\|p\|^4 + \frac{1}{2}\|p\|^2 + \frac{1}{16}\right) + o(\|\triangle p_2\|^2) \\ &= \frac{\|\triangle p_2\|^2}{(\|p\|^2 + 1/4)^2} + o(\|\triangle p_2\|^2). \end{aligned}$$

Then we can again compute the change of area formula as

$$\begin{aligned} |\det &\left(D_p T(x, p)\right)| \\ &= \lim_{\|\triangle p_1\|, \|\triangle p_2\| \to 0} \frac{\|T(x, p) - T(x, p + \triangle p_1)\|\|T(x, p) - T(x, p + \triangle p_2)\|}{\|\triangle p_1\| \|\triangle p_2\|} \\ &= \lim_{\|\triangle p_1\|, \|\triangle p_2\| \to 0} \frac{\|\triangle p_1\|\|\triangle p_2\|/(\|p\|^2 + 1/4)^2 + o\left(\|\triangle p_1\|\|\triangle p_2\|\right)}{\|\triangle p_1\| \|\triangle p_2\|} \\ &= \frac{1}{(\|p\|^2 + 1/4)^2}. \end{aligned}$$

Hence, the determinant of the mixed Hessian for the logarithmic cost is

$$\left|\det D_{xy}^2 c(x, y)\right| = \left(\|p\|^2 + 1/4\right)^2. \tag{A.2}$$

References

[1] S. Haker, L. Zhu, A. Tannenbaum, S. Angenent, Optimal mass transport for registration and warping, Int. J. Comput. Vis. 60 (3) (2004) 225–240, https://doi.org/10.1023/B:VISI.0000036836.66311.97.
[2] U. Frisch, S. Matarrese, R. Mohayaee, A. Sobolevski, A reconstruction of the initial conditions of the universe by optimal mass transportation, Nature 417 (2002), https://doi.org/10.1038/417260a.
[3] M.J.P. Cullen, J. Norbury, R.J. Purser, Generalised Lagrangian solutions for atmospheric and oceanic flows, SIAM J. Appl. Math. 51 (1) (1991) 20–31, https://doi.org/10.1137/0151002.
[4] G. Peyré, M. Cuturi, Computational optimal transport: with applications to data science, Found. Trends Mach. Learn. 11 (5–6) (2019) 355–607.
[5] B. Engquist, B.D. Froese, Application of the Wasserstein metric to seismic signals, Commun. Math. Sci. 12 (5) (2014) 979–988.
[6] J.-D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. 84 (3) (2000) 375–393.
[7] J.-D. Benamou, V. Duval, Minimal convex extensions and finite difference discretisation of the quadratic Monge-Kantorovich problem, Eur. J. Appl. Math. (2017) 1–38.
[8] J.-D. Benamou, B.D. Froese, A.M. Oberman, Numerical solution of the optimal transportation problem using the Monge-Ampère equation, J. Comput. Phys. 260 (2014) 107–126.
[9] C.J. Budd, J. Williams, Moving mesh generation using the parabolic Monge-Ampère equation, SIAM J. Sci. Comput. 31 (5) (2009) 3438–3465.
[10] B.D. Hamfeldt, Convergence framework for the second boundary value problem for the Monge-Ampère equation, SIAM J. Numer. Anal. 57 (2) (2019) 945–971.
[11] M. Lindsey, Y.A. Rubinstein, Optimal transport via a Monge-Ampère optimization problem, SIAM J. Math. Anal. 49 (4) (2017) 3073–3124.
[12] C.R. Prins, R. Beltman, J.H.M. ten Thije Boonkkamp, W.L. IJzerman, T.W. Tukker, A least-squares method for optimal transport using the Monge-Ampère equation, SIAM J. Sci. Comput. 37 (6) (2015) B937–B961.
[13] B. Schmitzer, A sparse multiscale algorithm for dense optimal transport, J. Math. Imaging Vis. 56 (2) (2016) 238–259.
[14] N.K. Yadav, J.H.M. ten Thije Boonkkamp, W.L. Ijzerman, A Monge-Ampère problem with non-quadratic cost function to compute freeform lens surfaces, J. Sci. Comput. 80 (1) (2019) 475–499.
[15] 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.
[16] H. Weller, P. Browne, C. Budd, M. Cullen, Mesh adaptation on the sphere using optimal transport and the numerical solution of a Monge-Ampère type equation, J. Comput. Phys. 308 (2016) 102–123.
[17] T. Glimm, V. Oliker, Optical design of single reflector systems and the Monge-Kantorovich mass transfer problem, J. Math. Sci. 117 (3) (2003) 4096–4108.
[18] X.-J. Wang, On the design of a reflector antenna II, Calc. Var. Partial Differ. Equ. 20 (3) (2004) 329–341.

[19] L.B. Romijn, J.H.M. ten Thije Boonkkamp, W.L. IJzerman, Inverse reflector design for a point source and far-field target, J. Comput. Phys. 408 (2020) 109283.
[20] B.F. Hamfeldt, A.G.R. Turnquist, A convergence framework for optimal transport on the sphere, arXiv preprint, arXiv:2103.05739, 2021.
[21] N.S. Trudinger, On the local theory of prescribed Jacobian equations, Discrete Contin. Dyn. Syst., Ser. A 34 (4) (2014) 1663.
[22] 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.
[23] M.G. Crandall, H. Ishii, P.-L. Lions, User's guide to viscosity solutions of second order partial differential equations, Bull. Am. Math. Soc. 27 (1) (1992) 1–67.
[24] G. Barles, P.E. Souganidis, Convergence of approximation schemes for fully nonlinear second order equations, Asymptot. Anal. 4 (1991) 271–283.
[25] 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.
[26] G. Loeper, On the regularity of solutions of optimal transportation problems, Acta Math. 202 (2009) 241–283.
[27] B.D. Froese, A.M. Oberman, Convergent finite difference solvers for viscosity solutions of the elliptic Monge-Ampère equation in dimensions two and higher, SIAM J. Numer. Anal. 49 (4) (2011) 1692–1714.
[28] B.D. Froese, Meshfree finite difference approximations for functions of the eigenvalues of the Hessian, Numer. Math. 138 (1) (2018) 75–99.
[29] H. Schaeffer, T.Y. Hou, An accelerated method for nonlinear elliptic PDE, J. Sci. Comput. 69 (2) (2016) 556–580.