Earth Mover’s Distances on Discrete Surfaces

Justin Solomon  
Stanford University

Raif Rustamov  
Stanford University

Leonidas Guibas  
Stanford University

Adrian Butscher  
Max Planck Institute for Informatics

Abstract

We introduce a novel method for computing the earth mover’s distance (EMD) between probability distributions on a discrete surface. Rather than using a large linear program with a quadratic number of variables, we apply the theory of optimal transportation and pass to a dual differential formulation with linear scaling. After discretization using finite elements (FEM) and development of an accompanying optimization method, we apply our new EMD to problems in graphics and geometry processing. In particular, we uncover a class of smooth distances on a surface transitioning from a purely spectral distance to the geodesic distance between points; these distances also can be extended to the volume inside and outside the surface. A number of additional applications of our machinery to geometry problems in graphics are presented.

CR Categories:  I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems; G.1.6 [Numerical Analysis]: Optimization—Convex programming

Keywords: Optimal transportation, Wasserstein metric, earth mover’s distance, finite elements, geometric median

Links: ✨DL ✉PDF

1 Introduction

A common task in geometry processing is the computation of various classes of distances between points on or inside a discrete surface. For example, many shape matching algorithms need clues about the relative positioning and orientations of features to be matched, which can be obtained from pairwise feature distances. It is desirable for such distances to be true metrics, intrinsic, globally shape-aware, smooth, and insensitive to noise and topology changes, without inducing considerable distortion on the underlying metric. In particular, the level sets of the distance function should be evenly spaced, in a visual sense, along the surface.

Existing approaches to defining and computing intrinsic distances do not satisfy all of these requirements. Despite their central place in classical differential geometry, geodesic distances have many shortcomings for computational applications, such as being sensitive to noise and topology and not being globally shape-aware, that is, not conforming to geometric features of the surface as distances increase [Lipman et al. 2010]. Spectral distances [Coifman et al. 2005; Fouss et al. 2007; Lipman et al. 2010] overcome these shortcomings but can be unintuitive with unevenly-spaced isocontours. The drawbacks of geodesic and spectral distances indicate that a hybrid approach is needed. The approximation of geodesic distance formulated in [Crane et al. 2013] is an important step in this direction. While these approximations are smoothed versions of the geodesic distance that are robust and globally shape-aware, they may not be symmetric or satisfy the triangle inequality.

In this paper, we introduce a novel hybrid approach for computing a variety of surface distances that have all of the desired properties. The key idea is to consider the more general problem of computing distances between probability distributions supported on the mesh. In fact, our goal is to compute the gold standard of distributional distances, the well-known earth mover’s distance (EMD). Once we have a means to compute the EMD between general distributions, we then consider various specializations that lead to the surface distances of this paper, including a new approach to computing geodesic distance.

The computation of the EMD on a surface could be performed using a brute force linear programming approach. This approach, however, not only is computationally infeasible on reasonably-sized meshes but also leads to a “chicken-and-egg problem,” since such a formulation requires precomputing all pairwise geodesic distances between mesh vertices. Therefore, one of our contributions is to make use of an alternative differential formulation of EMD that can be discretized using finite elements (FEM). Furthermore, a spectral expansion of the optimization variables reveals successive approximations to EMD that are fast to compute and are themselves distance metrics.

Our approach has several practical benefits. First, our family of distances is general and provides a principled and efficient way to compute distances between probability distributions supported on various types of features. Second, if we consider delta distributions centered on surface points, we obtain a family of distances that ranges from the geodesic distance (since the EMD between two delta distributions reduces to geodesic distance) to a novel spectral distance. This latter distance is perceptually convincing (Figure 1) despite having a simple formulation in terms of the Green’s function of the Laplace-Beltrami operator. Third, inspired by [Rustamov et al. 2009], we develop a means of extending our distances
Overview

After discussing existing approaches for distance computation (§2), we formulate a differential method for computing earth mover’s distances on continuous surfaces (§3.1) and propose theoretical properties suggesting its suitability for geometry processing (§3.2). After simplification, we propose a discretization for triangle meshes (§4); the discretization itself satisfies the triangle inequality for discrete distributions (§4.4) and can be optimized using a simple iterative algorithm (§4.5). We then return to the computation of pointwise distances on surfaces, showing how our new distributional distance can be applied to this problem (§5); we also extend our distance to a distance metric on all of $\mathbb{R}^3$, solving an open problem proposed in [Rustamov et al. 2009] (§6). Finally, we propose applications of our distances to other geometry problems in graphics (§7) and suggest directions for future work (§8).

2 Related Work

Distances

Existing approaches to defining and computing intrinsic distances broadly can be categorized as “primal” or “dual.” A primal approach operates on the surface mesh directly to obtain exact or approximate distances. For example, [Mitchell et al. 1987; Surazhsky et al. 2005] apply such an approach for finding exact geodesic distances (i.e. lengths of shortest paths constrained to lie on the mesh), while [Kimmel and Sethian 1998; Campen et al. 2013] approximate geodesic distances to achieve faster run times. As discussed in [Lipman et al. 2010], however, despite their connection to classical differential geometry, geodesics have a number of shortcomings for computational applications, such as being sensitive to noise and topology and not being globally shape-aware.

These shortcomings inspired the development of dual methods for distance computation. Dual distances lift the problem to an alternative space, such as the set of real-valued functions on the mesh, where relationships between function values are used as proxies for inferring distances on the underlying domain. The most popular dual distances are spectral distances, such as the diffusion [Coifman et al. 2005], commute time [Fouss et al. 2007], and biharmonic [Lipman et al. 2010] distances. These distances can be unintuitive, however, with isocontours that are unevenly-spaced along the surface. This artifact is a fundamental problem, because dual approaches achieve global shape-awareness and robustness by averaging over many paths whose structure can depend on the curvature and local diameter of the surface.

The drawbacks of completely primal or dual methods indicate that a hybrid approach integrating properties of both approaches may be called for. In [Crane et al. 2013] an approximation of geodesic distance was formulated by integrating the normalized gradient field of the heat kernel. This approximation gives smoothed versions of the geodesic distance parameterized by the time parameter of the heat kernel. While robust and globally shape-aware, these are not guaranteed to be true distance metrics. Another hybrid distance was proposed in [Panzo et al. 2013], where geodesics between sampled vertices are embedded in Euclidean space using multi-dimensional scaling (MDS). The embedding is interpolated to the entire mesh by solving a biharmonic equation, and Euclidean distances in the embedding space provide a distance measure on the entire mesh. Since it is generally impossible to embed geodesic distances exactly into Euclidean space, this approach is likely to give inconsistent results when run repeatedly.

Optimal Transportation

Our work draws principally from the theory of optimal transportation; relevant aspects of this theory are summarized in §3. The usual discretization of EMD solves a linear program with one variable for each pair of points on the domain. The complexity of this program thus scales quadratically with the number of points in the domain, and requires a matrix of pairwise distances. Assorted approximations of EMD truncate large distances [Pele and Werman 2009] or are specialized to discrete domains like graphs [Takano and Yamamoto 2010].

Whereas EMD is the one-Wasserstein distance, more attention has been paid to computation and theoretical understanding of two-Wasserstein distances, which employ squared geodesic distances. These have stronger regularity properties and also admit a differential formulation, but discretizing this formulation requires an additional time variable that multiplies the number of variables to optimize and leads to challenging variational problems, e.g. [Benamou and Brenier 2000]. Discrete instances of the two-Wasserstein distance in $\mathbb{R}^n$ can yield specialized computation techniques [Mercigot 2011; Gu et al. 2013], although extension to distributions on surfaces and other domains is challenging.

In graphics and vision, EMD and its optimal transportation counterparts have been applied to a variety of problems. EMD was first introduced to the vision community in [Rubner et al. 2000] and since has been used to compare histograms and other descriptors. More recently, [Bonneel et al. 2011] applies approximations of optimal transportation to interpolate between BRDFs, intensity histograms, and other simple distributions; similar problems are considered in [Bonneel et al. 2013] after defining the barycenter of a set of distributions with respect to approximated transportation distances. [de Goes et al. 2011; de Goes et al. 2012] compute transportation distances from two-dimensional point sets for application in shape processing and blue noise generation, while [Mullen et al. 2011] employs a similar formulation to triangulation problems. These distances also have been applied to geometry analysis [Lipman and Daubechies 2011; Lipman et al. 2013], spherical parameterization [Dominitz and Tannenbaum 2010], and matching [Mémoli 2011; Solomon et al. 2012]. None of these approaches, however, is able to compute EMDs or related distributional distances intrinsic to meshed geometry without aggressive approximation or restriction to a simpler domain.

Perhaps the closest construction to ours in the computer graphics literature is in [Solomon et al. 2013]. This paper uses the two-Wasserstein distance—whereas we consider the one-Wasserstein distance—to analyze maps from points on one surface to distributions on another. This paper is able to discretize only a linearization of transportation distances between similar distributions, which is sufficient for their map continuity energy; optimizing this distance is intractable in their formulation and hence their paper is applied to map analysis rather than computation. The idea of endowing geometric objects with probability measures and replacing Euclidean distances with Wasserstein distances also appears in the distance-to-measure work of [Chazal et al. 2010; Chazal et al. 2011]. In that work, however, the main goal is to define distance functions from points to point clouds that are robust to outliers, so that appropriate stability results can be proven for topological inference.

3 Distance Computation

3.1 Optimal Transportation

Suppose $M$ is a manifold with or without boundary. We can describe a mass distribution on $M$ using a measure. Omitting some technical detail, this is simply a positive, real-valued function $\mu$ acting on subsets of $M$, where we interpret the value $\mu(U)$ as the
amount of mass contained in $U \subseteq M$. Examples of measures are: (1) if $p : M \to \mathbb{R}_+$ is an integrable function where $\rho(x)$ gives the mass density at $x \in M$, then we obtain the measure $\mu(U) \equiv \int_U \rho(x) \, dx$; and (2) the Dirac mass at $p \in M$, defined by its action on subsets as $\delta_p(U) \equiv 1$ if $p \in U$ and $\delta_p(U) \equiv 0$ otherwise. A useful feature of measures is that one can integrate real-valued functions: if $\phi : M \to \mathbb{R}$ describes some property of the mass distribution as a function of position in $M$, then we interpret the integral $\int_M \phi(x) \, d\mu(x)$ (also denoted $\int_M \phi(x) \, d\rho(x)$) when $\mu$ has density $\rho$ as proportional to the average value of $\phi$ over the mass distribution. Finally, if the total mass of $\mu$ equals one, then we say that $\mu$ is a probability distribution. We denote the set of probability measures on $M$ as $\text{Prob}(M)$.

We refer the reader to [Folland 1999] for further information about measure theory.

We would like to compute a distance between any two probability distributions. That is, we seek a positive definite, real-valued function $\text{Prob}(M) \times \text{Prob}(M) \to \mathbb{R}_+$ satisfying these constraints. In the language of fluid dynamics, it can be thought of following convex optimization yields the EMD between admitting densities $x,y$. Discretizations of $M$ ally, if we are to use general optimization machinery to compute efficiently as possible, the following proposition is intuitively clear efficient.

Now let $\Pi(\mu_0, \mu_1)$ denote the set of transportation plans satisfying these constraints, and let $d(\cdot, \cdot)$ be the geodesic distance function in $M$. Then we define the $1$-Wasserstein distance as the optimal value

$$ W(\mu_0, \mu_1) \equiv \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{M \times M} d(x, y) \, d\pi(x, y). \quad (1) $$

Therefore $W(\mu_0, \mu_1)$ can be interpreted as the cost of the optimal plan transporting the mass of $\mu_0$ to that of $\mu_1$, when moving mass from $x$ to $y$ costs $d(x, y)$. We refer the reader to [Villani 2003] for further information concerning the theory of optimal transportation.

Discrete approximations of earth mover’s distance that are suitable for computer vision and graphics are difficult to obtain due to computational complexity. For instance, a discretization of $\pi \in \Pi(\mu_0, \mu_1)$ in (1) requires a quadratic number of variables, since $\pi(x, y)$ is a function of two positions $x, y \in M$. Additionally, if we are to use general optimization machinery to compute $W(\mu_0, \mu_1)$ we must be able to precompute or approximate $d(x, y)$ for any pair $x, y \in M$. This scaling, which appears in many known discretizations of $W$, is prohibitively expensive for large meshes.

A remarkable observation from the theory of optimal transportation provides a differential strategy for evaluating $W$ under suitable regularity. Suppose $M$ is a compact surface and $\mu_0, \mu_1 \in \text{Prob}(M)$ admit densities $\rho_0, \rho_1 : M \to \mathbb{R}_+$. Then, the optimal value of the following convex optimization yields the EMD between $\rho_0$ and $\rho_1$:

$$ W(\mu_0, \mu_1) \equiv \left\{ \begin{array}{ll} \inf_{J} \int_M \|J(x)\| \, dx \\ \text{s.t. } \nabla \cdot J(x) = \rho_1(x) - \rho_0(x) \\ J(x) \cdot n(x) = 0 \quad \forall x \in \partial M \end{array} \right. \quad (2) $$

This optimization computes a vector field $J$ on $M$ whose boundary $\partial M$ has normal $n(x)$ using a convex energy with linear constraints. In the language of fluid dynamics, it can be thought of as an Eulerian alternative to the Lagrangian formulation (1), i.e. points $x \in M$ watch probabilistic mass move past along flow lines of $J(x)$. Indeed, (2) first arose as the “Beckmann problem” in network flow [Beckmann 1955]. See [Samantambrogio 2013] for further analysis of this problem and its connection to optimal transportation, and see [Villani 2003, §1.2.3], [Feldman and McCann 2002], and [Samantambrogio 2009] for a broader discussion.

In this paper, we use (2) as a starting point for the computation of a discrete approximation of the $1$-Wasserstein distance between two probability distributions on $M$. With only one unknown $J(x)$ per point $x \in M$, this optimization scales linearly with the size of the mesh rather than the quadratic scaling of (1).

### 3.2 Properties of the 1-Wasserstein Distance

Given its definition via geodesics distances in (1), it comes as no surprise that the $1$-Wasserstein distance $W$ is intricately linked with the metric structure of $M$ even if it is computed using differential techniques. Here we state some properties relevant to our target applications in geometry processing and graphics; proofs can be found in the Appendix.

The Beckmann problem recasts the transportation problem (1) in Eulerian language as finding the direction of steady-state flow of mass from $\rho_0$ to $\rho_1$. The vector field $J$ can be thought of the velocity of this flow; examples of $J$ computed using the discrete method in §4 are shown in Figure 2. Since an optimal flow moves mass as efficiently as possible, the following proposition is intuitively clear and follows from the first-order optimality conditions for (2):

**Proposition 1.** Let $J$ solve the optimization (2) with given densities $\rho_0, \rho_1$. Then, flow lines of $J$ are geodesics on $M$.

A feature of the $1$-Wasserstein distance distinguishing it from $p$-Wasserstein distances with $p > 1$ is that its optimal transportation plans are not unique [Villani 2003, §2.4.6]. That is, it is known that mass is transported along geodesics, but not how far a particle of mass travels along any given geodesic. One optimal plan obtained from $J(x)$ is the Dacorogna-Moser construction [Villani 2003, Chapter 5]. In this construction, we let $J$ solve (2) and define $\rho_t \equiv (1 - t)\rho_0 + t\rho_1$. Now we consider the flow $z : [0, 1] \times M \to M$ of the ordinary differential equation (ODE):

$$ \dot{z} = -\frac{J(z)}{\rho_t(z)}. \quad (3) $$

So, $z$ produces a curve $t \mapsto z(t, x_0) \in M$ satisfying the ODE with initial conditions $z(0, x_0) = x_0$. Moreover, since the velocity of this curve is proportional to $J$, it is also a geodesic. The Dacorogna-Moser plan takes the mass at each $x_0 \in M$ and moves it to $z(1, x_0)$.

Next, by definition the Eulerian velocity of the Dacorogna-Moser flow is $V_t(x) = -J(x)/\rho_t(x)$. Consequently, the transport equation $\frac{\partial}{\partial t} + \nabla \cdot (\rho_t V_t) = 0$ satisfied by $(\rho_t, V_t)$ shows $\nabla \cdot J = \rho_1 - \rho_0$. This tells us that if $\rho_0$ is advected under the flow $z$, the resulting time-dependent family of densities is $\rho_t$. 

![Figure 2: Vector fields $J$ transporting mass to a distribution concentrated on the nose of a horse from a distribution on (a) one of its hooves, (b) all four of its hooves, and (c) its tail.](image-url)
Finally, the following result shows that one can recover pointwise geodesic distances from the 1-Wasserstein distance in the special case of infinitely sharply peaked δ-distributions.

**Proposition 2.** Let δp be a delta distribution centered at p ∈ M, and let χQ be the uniform distribution supported on a subset Q ⊆ M. Let J be the solution of the optimization problem (2). Then,

1. The flow lines of J are geodesics from p to all points of Q.
2. As Q → {q} and χQ → δq appropriately, then W(δp, χQ) converges to the geodesic distance between p and q.

4 Simplification and Discretization

The formulation of EMD in (2) is amenable to discretization on triangle meshes, commonly encountered in graphics and geometry processing. In this section we propose a discretization using finite elements (FEM) that admits straightforward optimization.

4.1 Vector Field Decomposition

The Helmholtz-Hodge decomposition shows that any vector field J on M can be written as [Schwarz 1995; Polthier and Preiß 2003]:

\[ J(x) = \nabla f(x) + \mathcal{R} \cdot \nabla g(x) + h(x) \]

where \( \mathcal{R} \) is the linear operator that rotates a vector 90° clockwise in the tangent plane. The “gradient part” of J is the vector field \( \nabla f \), the “curl part” of J is the vector field \( \mathcal{R} \cdot \nabla g \), and the “harmonic part” of J is the vector field h satisfying \( \nabla \cdot h(x) = 0 \) and \( \nabla \times h(x) = 0 \). In case \( \partial M \neq \emptyset \), we must additionally impose Neumann boundary conditions on J. This boundary condition reflects the fact that shortest-path curves cannot leave the surface, so when they reach boundary they must become tangential.

Substituting this decomposition into (2) and using the fact that \( \nabla \cdot (\mathcal{R} \cdot \nabla g) = \nabla \cdot h = 0 \), we obtain the equivalent optimization:

\[
\mathcal{W}(\mu_0, \mu_1) = \begin{cases} 
\inf_{f,g,h} \int_M \| \nabla f(x) + \mathcal{R} \cdot \nabla g(x) + h(x) \| \, dx \\
\text{s.t.} \, \Delta f(x) = \rho_1(x) - \rho_0(x), \\
g(x) = 0 \text{ and } \partial \rho(x) = 0 \forall x \in \partial M; \\
\nabla \cdot h(x) = 0 \text{ and } \nabla \times h(x) = 0.
\end{cases}
\]

This form shows that f is determined independently of the optimization by solving \( \Delta f = \rho_1 - \rho_0 \) with Neumann boundary conditions; this equation has a solution since \( \rho_0 \) and \( \rho_1 \) integrate to 1. We therefore compute \( \mathcal{W} \) in two steps, one for finding f and one for finding g and h; moreover, eliminating f from the optimization leaves an essentially unconstrained optimization for g and h.

4.2 Spectral Reduction

We further simplify (4) using the spectral decomposition of the Laplacian \( \Delta \). First, we obtain a basis for functions on M by solving the eigenvalue problem \( \Delta \phi_i = \lambda_i \phi_i \) with Dirichlet boundary conditions \( \phi_i|_{\partial M} = 0 \). The gradients and rotated vectors of these functions comprise a basis for gradient and curl fields of M. Additionally, the set of harmonic vector fields on M admits a basis with dimension equal to twice the genus of M [Tong et al. 2006].

Denoting a combined basis for curl fields and for harmonic fields as \( \psi_1, \psi_2, \psi_3, \ldots \), we can write the unknown vector field as \( \mathcal{R} \cdot \nabla g + h = \sum c_i \psi_i \) where \( c_i \) are unknown coefficients. After solving \( \Delta f = \rho_1 - \rho_0 \) to precompute the vector field \( v = \nabla f \), (4) can be recast as the unconstrained optimization:

\[
\mathcal{W}(\mu_0, \mu_1) = \inf_{\{c_i\}} \int_M \| v(x) + \sum_i c_i \psi_i(x) \| \, dx.
\]

This objective is convex in the unknowns \( c_i \). Boundary conditions are not needed because they have been incorporated into the \( \psi_i \)'s.

4.3 Discretization via Finite Elements (FEM)

Using triangle mesh geometry to discretize M, we express scalar functions \( f : M \to \mathbb{R} \) with one value per vertex interpolated to faces using piecewise linear “hat” functions. Vector fields are piecewise constant per face, allowing for a gradient operator \( \nabla \) taking functions on the vertices to vector fields on the faces.

We solve the Poisson equation \( \Delta f = \rho_1 - \rho_0 \) with Neumann boundary conditions for f using a first-order finite elements approach as in [Sayas 2008]. This sparse linear solve can be carried out at interactive rates without the need for spectral approximation. Then, we compute the curl and harmonic components of f. For the curl vector fields, we examine two options for choosing a basis as above, trading off between speed and quality. The most accurate solutions are obtained simply by writing \( g \) with one value per vertex. Alternatively, we can improve timings with some cost in accuracy by writing \( g \) in a truncated basis of low-frequency eigenvectors of the Laplacian matrix \( \Delta \). We use a method like [Tong et al. 2006] to compute a basis for harmonic vector fields.

In our discretization, (5) becomes the following optimization:

\[
\inf_{\{a_t\}} \sum_{t \in T} a_t \| v_t + \sum_i c_i \psi_i |_t \|.
\]

where T is the set of triangles in M, each triangle \( t \in T \) has area \( a_t \); \( \psi_i |_t \) is the value of the basis element \( \psi_i \) on triangle \( t \), and \( v_t \) is the gradient of the piecewise-linear \( f \) defined above.

If we use a truncated eigenbasis for \( g \), we are only approximating the distance \( \mathcal{W} \). Effectively this constrains certain coefficients \( c_i \) of the expansion in (5) to zero, and thus these approximations overestimate \( \mathcal{W} \). Figure 3 illustrates convergence as the number of eigenfunctions is increased; even a small spectral basis provides a strong approximation of \( \mathcal{W} \).

4.4 Properties of the Discretization

Since we discretized (5) we can expect its properties to hold approximately for the discretization. We can, however, prove that one important property of the discretization holds exactly, even with spectral approximation:

**Proposition 3.** Minima of (6) satisfy the triangle inequality for discrete probability distributions represented using one value per vertex on M, even if the bases for curl and harmonic vector fields are truncated.

Thus, our approximations of \( \mathcal{W} \) are in fact distances in themselves.

4.5 Optimization

We derive an algebraic form for (6) by assembling the coefficients \( c_i \) into a vector \( c \) and the vectors \( \psi_i |_t \) for a given triangle \( t \) into the columns of a matrix \( A_t \in \mathbb{R}^{\times k} \). After defining \( w_t \equiv a_t v_t \) and \( B_t \equiv a_t A_t \), (6) becomes the minimization problem

\[
\inf_{\{c\}} \sum_t \| B_t c + w_t \|.
\]
notes the least-squares solution outlined in [Li 1998]; in challenging test cases the least-squares to be pre-factored for and solves an identical linear system in each iteration, allowing it ers (ADMM) [Boyd et al. 2011]; this new approach (derived in the optimization literature and can be solved using a vari-geometric median” or “continuous location” problem, appears In this form, our optimization problem attempts to minimize a function for i = 1, 2, 3, . . . Iterate until convergence zt ← Btc + wt − ut Merge and solves an identical linear system in each iteration, allowing it to be pre-factored for all EMD computations on a surface. Figure 3: (a) Two distributions p0, p1 on a sphere colored yellow and blue; (b) the Hodge decomposition of the vector field J taking p0 to p1; (c) approximations of J with more and more curl basis functions (basis size on upper right of each sphere); (d) EMD between p0 and p1 as a function of basis size.

function ADMM-WASSERSTEIN(ρ0, ρ1) 
\[ \triangledown \] ρ0, ρ1 have one value per vertex
\[ \triangledown \] Concatenate Bt’s vertically to obtain B
\[ \triangledown \] Solve for gradient part
\[ \triangledown \] Compute gradient vector field
for i ← 1, 2, 3, . . . Iterate until convergence
zt ← Btc + wt − ut Merge
Jt ← αtzt Merge and solves an identical linear system in each iteration, allowing it to be pre-factored for all EMD computations on a surface.

return Jt \( \forall t \in T \)

function WEISZFELD-WASSERSTEIN(p0, p1) 
\[ \triangledown \] p0, p1 have one value per vertex
\[ \triangledown \] Concatenate Bt’s vertically to obtain B
\[ \triangledown \] Solve for gradient part
\[ \triangledown \] Compute gradient vector field
\[ \triangledown \] Initialize vector field to zero
for i ← 1, 2, 3, . . . Iterate until convergence
rt ← Btc + wt Merge
R ← rt 1 Merge and solves an identical linear system in each iteration, allowing it to be pre-factored for all EMD computations on a surface.

\[ \alpha_t \left\{ \begin{array}{ll} 1 - \frac{1}{\beta \|z_t\|} & \beta \|z_t\| > 1 \\ 0 & \text{otherwise} \end{array} \right. \]
Jt ← αtzt Merge
\[ \triangledown \] Update coefficients; can pre-factor
\[ \rho_t \left\{ \begin{array}{ll} \sum_i B_t(i) (B_t(i)^{-1} \left( \sum_i B_t(i)^{-1} (\frac{w_i + J_t - Btc}{w_i} - w_i) \right) \right) \]
yt ← yt + β(Jt − Btc − wt Merge and solves an identical linear system in each iteration, allowing it to be pre-factored for all EMD computations on a surface.

\[ \beta \rho_1 \rho_1 \]
\[ \triangledown \] Update dual
return Jt \( \forall t \in T \)

Figure 4: Weiszfeld algorithm for optimizing \( W \), using the steps outlined in [Li 1998]; in challenging test cases the least-squares solve can be regularized slightly for numerical stability. \( A^+ R \) denotes the least-squares solution X to the system AX \( \approx R \).

Figure 5: ADMM algorithm for optimizing W, derived in the supplemental document, with parameter \( \beta > 0 \).

5 Pointwise Distances

In this section we consider the problem of computing intrinsic distances between points on surface meshes. Using our machinery, we first introduce a family of pointwise distance metrics \( d_{\nu}^{k}(\cdot, \cdot) \), with \( k = 0, \ldots, n_{\text{vert}} \), and state their theoretical properties. Next, we investigate the practical properties of these distances, including their behavior on realistic meshes, qualitative comparison to commonly used distances, and empirical sensitivity to mesh perturbations.

To define the distance between two mesh points \( p, q \in M \), we consider two distributions \( \delta_p, \delta_q \in \text{Prob}(M) \) that have mass only at \( p \) and \( q \), resp.; discretely, these distributions are nonzero only at individual vertices. Then, we define a distance metric on \( M \) as \( d_{W}(p, q) \equiv W(\delta_p, \delta_q) \). We can compute this distance using spectral approximation as in Section 4 with basis size \( k \) and denote this approximate distance by \( d_{W}^{k}(\cdot, \cdot) \). Note that \( k = 0 \) corresponds to the case when the curl and harmonic terms are removed altogether.

Properties of our family of pointwise distances follow directly from our discussion in previous sections. Both in the discrete and the continuous cases, \( d_{W}^{k}(\cdot, \cdot) \) is a true distance metric for all values of \( k \). Also, given any two points \( p, q \in M \), we have

\[ d_{W}^{k}(p, q) \geq d_{W}(p, q) \geq d_{W}^{\text{vert}}(p, q) = d_{\nu}(p, q), \]

where \( d_{\nu}(\cdot, \cdot) \) is a discretization of geodesic distance. This final distance may not coincide exactly with the discrete geodesic distance along triangle faces but still satisfies symmetry, the triangle equality, and a discretization of the eikonal equation simultaneously.

The initial member of our family, \( d_{W}^{0}(\cdot, \cdot) \), has a particularly simple form suitable for efficient implementation. This distance can be computed as \( \int_{M} \| \nabla f(x) \| \, dx \), where \( f \) satisfies \( \nabla f = \delta_p - \delta_q \). It follows that \( f(x) = G(x, p) - G(x, q) \), where \( G(\cdot, \cdot) \) is the Green’s function of Laplace-Beltrami operator, establishing that:

\[ d_{W}^{0}(p, q) = \int_{M} \| \nabla s(G(x, p) - G(x, q)) \| \, dx \]

We can compare \( d_{W}^{0} \) to a state-of-art spectral distance, the biharmonic distance of [Lipman et al. 2010], which can be written (see the Appendix):

\[ d_{\nu}(p, q) = \left( \int_{M} (G(x, p) - G(x, q))^2 \, dx \right)^{\frac{1}{2}}. \]
Despite the resemblance, there are fundamental differences: In addition to taking gradients, our distance is based on an $\ell_1$ rather than $\ell_2$ norm. It is a classical result that $\ell_2$-norms have smaller embedding capacity than $\ell_1$ in that a point set that can be embedded isometrically into $\ell_2$ can also be embedded into $\ell_1$, but not vice versa [Deza and Laurent 2009]. In the context of geodesic distances on manifolds, the main obstruction is that non-unique geodesics cannot be supported by Euclidean distance. Indeed, every point $x$ on a shortest geodesic connecting points $p$ and $q$ satisfies $d_g(p, x) + d_g(x, q) = d_g(p, q)$, and under isometric embedding into Euclidean space this means that the image of $x$ lies on the segment connecting the images of $p$ and $q$ under the embedding. For example, geodesic distances on the sphere can be embedded isometrically in $\ell_1$ (c.f. [Deza and Laurent 2009]) but not $\ell_2$, since considering all the geodesics connecting the two poles it follows that the entire sphere must be mapped to a straight segment. This helps explain why existing spectral distances, largely based on $\ell_2$ norms, have nonintuitive disparately spaced isocontours. Contrastingly, we will see that $d_{bV}$ has isocontours that are relatively evenly spaced; we attribute this property to the larger embedding capacity of $\ell_1$-norm and to the fact that it has much in common with the larger optimization for $d_{bV}$.

**Experiments**  
Figure 6 shows examples of $d_{bV}^k$ for increasing $k$ and comparison with biharmonic distances $d_b$ [Lipman et al. 2010]. Some anisotropic behavior is specific to these “primitive” shapes, which have spectra with repeated eigenvalues that cannot be grouped in multiples of five; this phenomenon is unlikely on general shapes.

A few examples in Figure 7 typify the advantages of our new distances. On the boy model, we can see that biharmonic distances are strongly anisotropic in the horizontal direction and unevenly spaced away from the center point; on the other hand, geodesic distances on the back of the same model have numerous artifacts due to a lack of differentiability. The bearing model also shows similar anisotropy for $d_b$, while as $k$ increases our distances are able to capture level sets on the cap of the mesh. Our distances also are stable in the hair of the bust model, maintaining a reasonable distribution and smoothness despite high-frequency changes in geometry.

Figure 8 demonstrates the stability of $d_{bV}^0$ and $d_{bV}^{100}$ to common geometric perturbations. Figure 8(a) shows the insensitivity of these distances to per-vertex noise. Here the addition of Gaussian noise to the mesh leads to little change in the distance as evidenced by the coloring and the isocontours. Figure 8(b) confirms the theoretical isometry invariance property of our distance—the isocontours and coloring are in near correspondence between the armadillo model and its nearly isometric deformation. Finally, Figure 8(c) shows insensitivity to tessellation; the distance remains almost unchanged as the mesh is refined considerably.

Figure 9 compares our technique to [Crane et al. 2013]. Metric properties hold for our distances at all levels of spectral truncation even after discretization, while their smoothed geodesics at larger and larger diffusion times no longer benefit from an infinitesimal relationship with geodesic distances. This deviation can cause the triangle inequality to fail, as shown in red. Their smoothed distances also are not symmetric, that is, they may not satisfy $d(a, b) = d(b, a)$. While averaging forward and backward distances repairs this issue, it comes at the cost of considerably slower computation for tasks like finding the distance from a single source to all targets, replacing a single linear solve with one for each target.

**Computation time**  
Table 1 reports time to compute single-source distances, including geodesic and biharmonic, for a variety of meshes on a 2.40GHz Intel Xeon processor with 23.5GB RAM. The implementation is done in MATLAB, using the ADMM optimization in Figure 5. Our new spectral distance $d_{bV}$ is efficient to compute by factorizing the Laplacian and performs similarly to the
biharmonic distance; in fact, in this test $d_{W}^{0}$ outperforms $d_{W}$ considerably on larger meshes because it requires Green’s functions of the Laplace-Beltrami operator rather than the denser bilaplacian operator. Computing smoothed geodesic functions $d_{W}^{0}$ introduces computational cost scaling with the number of eigenfunctions.

Table 2 compares timing of computing all-pairs distances between a subsample of points on assorted meshes using $d_{W}^{0}$, $d_{W}$, and $d_{W}^{100}$. As in Table 1, the linear solve step for computing $d_{W}^{0}$ takes less time than that for $d_{W}$; integrating the derivative of the Laplace-Beltrami Green’s function, however, requires an additional iteration over the faces of the mesh, adding computation time for $d_{W}^{100}$ in this test. Here, $d_{W}$ and $d_{W}^{0}$ compute one Green’s function per source point and then only find pairwise distances between the prescribed points; no such optimization is available for $d_{W}$, which must compute distances to all vertices from each source point.

6 Volumetric Distances

The problem of computing volumetric distances respecting a given boundary mesh arises in a number of applications, e.g. path planning. Since the straightforward approach of computing shortest paths within a 3D polyhedron is NP-hard [Canny and Reif 1987], previous work introduced an approach based on interpolating a prescribed distance on the boundary mesh to the surrounding space inside and outside the shape [Rustamov et al. 2009]. While this approach is efficient, it requires an MDS-like embedding of prescribed pairwise distances and hence cannot exactly interpolate the geodesic distance.

Here, we show how our machinery can be used to obtain a volumetric distance reproducing geodesic distance when restricted to the boundary mesh. Like [Rustamov et al. 2009], we use barycentric coordinates but in a fundamentally different way—by considering them as distributions and computing EMDs between them.

For a given point $x$ in the interior of $M$, its barycentric coordinates with respect to a mesh $M$ with vertices $v_{i}, i = 1, \ldots, n_{\text{vert}}$, are weights $w_{i}(x), i = 1, \ldots, n_{\text{vert}}$. We recall three properties of these weights: the Lagrange property $w_{i}(v_{j}) = \delta_{ij}$ (the Kronecker delta); the partition of unity property $\sum_{i} w_{i}(x) = 1$ with $w_{i}(x) \geq 0$; and the linear precision property $\sum_{i} w_{i}(x)v_{i} = x$.

To compute our volumetric distance between $p$ and $q$, we consider their barycentric coordinates $\{w_{i}(p)\}_{i=1}^{n_{\text{vert}}}$ and $\{w_{i}(q)\}_{i=1}^{n_{\text{vert}}}$ as distributions $\mu_{p}$, $\mu_{q}$ in $\text{Prob}(M)$; this is possible due to the partition of unity property. Then, our distance is defined as $d_{W}(p, q) \equiv \mathcal{W}(\mu_{p}, \mu_{q})$; as before, this is a true distance metric. If $d_{W}(p, q) = 0$, then $\mu_{p} = \mu_{q}$, and so $p = q$, because by linear precision property, barycentric coordinates determine the point uniquely.

Our volumetric distance satisfies all of the relevant properties listed in [Rustamov et al. 2009]. If $W$ is computed without approximation, $d_{W}(p, q)$ reduces to geodesic distance when $p, q \in M$; indeed, due to the Lagrange property, in this case $\mu_{p} = \delta_{\mathcal{B}}, \mu_{q} = \delta_{\mathcal{B}}$, and we are back in the setting of previous section. Furthermore, the following maximum principle holds: if $p$ and $q$ have non-negative barycentric coordinates then the volumetric distance between $p$ and $q$ is no more than the geodesic diameter of the boundary mesh, i.e. $d_{W}(p, q) \leq \max_{x,y \in \partial M} d_{g}(x, y)$; this bound follows directly from (1) by upper-bounding $d(x, y)$. Also, we can show that $d_{W}$ is bounded below by Euclidean distance, that is $d_{W}(p, q) \geq ||p - q||$; the proof of this property is provided in the Appendix.

Our differential definition (2) of $d_{W}$ continues to be a distance metric when we allow $\mu_{p}$ and/or $\mu_{q}$ to have negative values in its density function, despite the weaker connection to the theory of optimal transportation. In fact, even our proof of the upper bound $d_{W}(p, q) \geq ||p - q||$ remains valid. This relaxation allows us to consider any choice of $p, q \in \mathbb{R}^{3}$ rather than restricting to the interior or convex hull of $M$, even if coordinates become negative.

Experiments In our experiments, we use mean value coordinates [Ju et al. 2005]; these coordinates can become negative both in the interior and exterior of $M$, but as noted above this departure from $\text{Prob}(M)$ does not raise any issues.

Figure 10 shows examples of this distance function in the space around a surface mesh. As before, we select a single source point and then compute the distance to other points in the volume near the object. We visualize these distances on two orthogonal planes using the same color coding as in the previous section.
Given an example, the distributions \( \rho_0 \) and \( \rho_1 \) are concentrated at two different points on the maze \( M \), which has two possible exits. The optimization \( (7) \) matches the two paths to their closest exits; this can be seen in the vector field \( J \). We also show the decomposition \( J = \nabla f + \nabla g \). Here, the gradient part \( \nabla f \) encodes large-scale motions in the maze while the rotational part \( \nabla g \) helps mass round corners in the maze efficiently; the quality of \( f \) alone reflects a connection to path planning algorithms using harmonic functions, e.g. [Connolly et al. 1990].

**Path planning** We can incorporate the convex energy for \( W \) into larger optimizations to formulate applications of our distances to more complex problems. Although the methods in Figures 4 and 5 no longer apply directly to the problems below, they are all convex and can be optimized using interior point methods.

As an initial example, the distributions \( \rho_0 \) or \( \rho_1 \) can be promoted to optimization variables to solve path planning problems. For example, suppose \( M \) is a mesh of a floor plan and that \( \rho_0 \) approximates a distribution of occupants in different parts of \( M \). To find the most efficient way to move all the occupants to a restricted subset of points \( S \subseteq M \), we can solve the optimization:

\[
\begin{align*}
\inf_{\rho_1} W(\rho_0, \rho_1) \\
\text{s.t. } \rho_1(x) = 0 \forall x \notin S \\
\rho_1(x) \geq 0 \forall x \in M \\
\int_M \rho_1(x) \, dx = 1
\end{align*}
\]

Adding a small multiple of \( \int_M \rho_1(x)^2 \, dx \) has a regularizing effect on \( \rho_1 \) when smoothness is desired.

Figure 12 shows this optimization applied to solving a maze. In this example, \( \rho_0 \) is concentrated at two different points on the maze \( M \), which has two possible exits. The optimization \( (7) \) matches the two points to their closest exits; this can be seen in the vector field \( J \), which traces a path from the starting points to their targets. We also show the decomposition \( J = \nabla f + \nabla g \). Here, the gradient part \( \nabla f \) encodes large-scale motions in the maze while the rotational part \( \nabla g \) helps mass round corners in the maze efficiently; the quality of \( f \) alone reflects a connection to path planning algorithms using harmonic functions, e.g. [Connolly et al. 1990].

**Fuzzy geodesics** Recall our intuition that the vector field \( J(x) \) is large at points \( x \) that see mass move past as \( \rho_0 \) advects toward \( \rho_1 \) according to the optimal matching. Suppose \( \rho_0 \) and \( \rho_1 \) are concentrated near two points \( p_0, p_1 \in M \), resp. Then, \( J(x) \) is large near geodesic curves between \( p_0 \) and \( p_1 \).

Inspired by [Sun et al. 2010], the norm \( \|J(\cdot)\| : M \to \mathbb{R} \) provides a “fuzzy geodesic” function related to the likelihood that a geodesic connecting points in the support of \( \rho_0 \) to points in the support of \( \rho_1 \) should pass through \( x \in M \). In particular, we can put small Gaussians around \( p_0 \) and \( p_1 \) and record \( \|J(\cdot)\| \) resulting from optimizing (2); an example is shown in Figure 13.

**Distance to features** There are many ways to use our framework to formulate distances that are aware of features rather than points. Most directly, to compute the distance from \( p \in M \) to \( S \subseteq M \), we can incorporate the convex energy for \( W \) via gradient descent on \( d_{W^\star}(|\cdot|, q) \) starting at \( p \). Since our distance is fast to compute, its gradient can be computed numerically in relatively little time. Figure 11 shows examples of paths from such a process, stepped using simple forward Euler integration. Remarkably, the time. Figure 11 shows examples of paths from such a process, stepped using simple forward Euler integration. Remarkably, the paths are tuned to the geometry of the boundary mesh and connect points in the interior without crossing \( M \).

**7 Additional Applications**

Here we suggest some less obvious applications of \( W \) to assorted geometry problems. These applications demonstrate the stability of our approach and suggest additional classes of problems for which it can be a valuable tool.
we can solve (7) with $\rho_0$ concentrated at $p$ and $\rho_1$ restricted to have support on $S$. A more efficient alternative, however, is suggested in the proof of Proposition 1. When computing $\mathcal{W}$, the Lagrange multiplier $\lambda : M \to \mathbb{R}$ for the $\nabla \cdot J = \rho_1 - \rho_0$ constraint satisfies the eikonal equation and hence is a geodesic distance. Thus, we can compute a geodesic function $\lambda$ that is aware of $S$ by computing $\mathcal{W}$ between the uniform distribution on $M$ and a distribution concentrated on the feature of interest and using the dual multipliers. Figure 14 shows this dual variable distance for computing distances to multiple points on a surface and distances to a curve.

Anisotropic distances As suggested in [Santambrogio 2013], the integral $\int_M \|J(x)\| \, dx$ from (2) can be replaced with a more general integral $\int_M \|A(x) \cdot J(x)\| \, dx$ to yield anisotropic transportation distances by modifying the metric of $M$. Figure 15 shows two examples in which the function $A$ is a nonnegative scalar guiding shortest paths to favorable areas or avoiding obstacles. In particular, inspired by [Lia et al. 2010] we are able to compute distances along a brain model that favor motion along ridges by weighting $J$ using mean curvature. More generally, matrix-valued $A$ can be used favor diffusion in a single direction; we leave consideration of the design of $A$ to future work.

Barycenters [Agueh and Carlier 2011] suggests minimizing a sum of transportation distances to find the barycenter of a set of distributions on a surface. Rather than resorting to approximations, e.g. in [Bonneel et al. 2011; Bonneel et al. 2013], our formulation of $\mathcal{W}$ allows us to solve this problem directly on a surface mesh.

Suppose we are given a set $\{\rho_1, \rho_2, \ldots, \rho_k\} \subset \text{Prob}(M)$. We can find a barycenter of these distributions by solving the optimization

$$\inf_{\rho} \sum_{i=1}^k \mathcal{W}(\rho_i, \rho)^2$$

s.t. $\rho(x) \geq 0 \forall x \in M, \quad \int_M \rho(x) \, dx = 1$ (8)

The distances $\mathcal{W}(\rho_i, \rho)$ are squared to imitate the units of the classical barycenter problem between points on $M$.

Figure 16 shows two applications of this optimization. In the first example, the barycenter of six distributions concentrated on the fingers and side of a hand model is centered at the upper palm; this output accurately represents points equally close to those favored by the six distributions. In the second example, a pointwise barycen-
adjustments to methods for computing. 

Figure 14: Distances computed from distributions (a) on the ears and tail of a pig model and (b) from the collar of a human model.

If the experiments in Figures 7, 9, and others show that our approximations into various geometric optimizations yields meaningful, and more accurate versions are easily evaluated using somewhat more involved flow techniques [Villani 2003]. While some specialized numerical methods exist for this case, our FEM discretization and spectral approximations may provide some insight into these problems and their applications in graphics.

We have focused on solving the differential EMD optimization problem (2) on triangle meshes, since they are the most common structures approached in geometry processing, but our formulation largely is general and could be applied to other structures admitting differential operators. For example, [Lai et al. 2013; Liang and Zhao 2013] and others present divergence, gradient, and Laplacian operators acting directly on point clouds and hence easily can be substituted into our formulation. More abstractly, [Jiang et al. 2011] and others adapt Helmholtz-Hodge structure to functions on graphs, revealing potential machine learning applications of our work.

Our distances also could be extended in various ways. The EMD is the p = 1 member of the class of p-Wasserstein distances between distributions. While in this paper we take advantage of the simple structure of the p = 1 case, Wasserstein distances with p > 1 have stronger regularity properties and can be evaluated using somewhat more involved flow techniques [Villani 2003]. While some specialized numerical methods exist for this case, our FEM discretization and spectral approximations may provide some insight into these problems and their applications in graphics.

We find empirically that the barycenter of a set of delta distributions is strongly peaked about a single point but defer a discussion of theoretical sharpness properties future work. In this case, however, the optimal optimization objective is exactly the sum of squared geodesic distances from the barycenter to each of the input points. In this way, this strategy reveals an alternative to [Panuzzo et al. 2013] for averaging sets of points on a surface.

8 Discussion and Conclusion

It is possible to envision many applications and extensions of the approaches to distance computation in this paper. As we have seen in many instances above, incorporating EMD and its spectral approximations into various geometric optimizations yields meaningful intrinsic information about the surface that easily and efficiently can be incorporated into machinery for larger problems.

While the experiments in Figures 7, 9, and others show that our technique has desirable properties compared to fast approximations, its runtime is limited thanks to the iterative optimization. Adjustments to methods for computing $d_{W^\kappa}$ could yield gains in efficiency. Approximating the least-squares solves for the Weiszfeld algorithm in Figure 4 may improve timing but more care is needed to guarantee convergence. The ADMM parameter $\beta$ in Figure 5 could be adjusted automatically, e.g. via [He et al. 2000], and alternative methods for geometric median problems may require fewer iterations [Li 1998; Qi et al. 2002; Zhou et al. 2003; Pan and Jiang 2008]. More generally, the optimization (2) is a second-order cone program, which can be minimized using commercial solvers.

Even without these extensions, our distances stand alone as practical tools for geometry processing. Even the lowest-order approximations of our distances are stable, smooth, and geometrically meaningful, and more accurate versions are easily evaluated using the proposed iterative method. Given the innumerable existing uses of EMD, spectral and geodesic distances, and volumetric distance, we are confident that differential earth mover’s distances will be a useful and straightforward alternative for characterizing the intrinsic geometry and relationships between features of surfaces.

Acknowledgments

The authors thank Nick Alger for discussions about ADMM, Andy Nguyen for discussions about path planning in §7, and Keenan Crane for providing an implementation of [Crane et al. 2013]. The authors gratefully acknowledge the support of NSF grant 1161480, ONR MURI N00014-13-1-0341, a Google research award, the Max Planck Center for Visual Computing and Communications, the Hertz Foundation Fellowship, and the NSF GRF program.

References


Appendix

Appendix

Proof of Proposition 1  If we let $\lambda : M \to \mathbb{R}$ be a Lagrange multiplier function for the divergence constraint, then the Lagrangian of (2) is given by:

$$L(J, \lambda) = \int_M \left( \|J(x)\| + \lambda(x) (\nabla \cdot J(x) + \rho_0(x) - \rho_1(x)) \right) dx$$

Let $J$ be a critical point of the variational problem (2) and introduce a variation vector field $\delta J$. Assume $\delta J|_{\partial M} \equiv 0$ so that the Neumann boundary condition is maintained. Also assume $\delta J(x) = 0$ whenever $J(x) = 0$. Taking the variation of the Lagrangian in the $\delta J$ direction yields:

$$0 = \frac{d}{d\varepsilon} L(J + \varepsilon \delta J, \lambda) \bigg|_{\varepsilon=0} = \int_M \delta J(x) \left[ \frac{J(x)}{\|J(x)\|} - \nabla \lambda(x) \right] dx.$$

Since this holds for all $\delta J$, we have shown $\nabla \lambda(x) = J(x)/\|J(x)\|$ whenever $J(x) \neq 0$. This shows that $\|\nabla \lambda(x)\| = 1$, and thus by the eikonal equation $\lambda$ is a geodesic distance function [Arnold 2003]; hence flows of $J(x)$ are either constant or geodesics.

Proof of Proposition 2  The first part is a consequence of Proposition 1 while the second part follows from the weak convergence properties of Wasserstein distances and the fact that Wasserstein distances between delta-distributions always reduce to geodesic distances since there is only one transport plan in $\Pi(\delta_p, \delta_q)$, namely the plan that assigns all the mass at $p$ to $q$.

Proof of Proposition 3  One can show directly that the quantity in the right hand side of (2) satisfies the properties of a distance without appealing to the equivalence of (2) with the 1-Wasserstein distance. That is, the right hand side of (2) is symmetric and non-negative, vanishing only if $J = 0$ or $\rho_i = \rho_j$. Moreover, the triangle inequality holds by the linearity of the divergence operator (so if $J_i$ satisfies $\nabla \cdot J_i = \rho_i - \rho_j$ then $\nabla \cdot (J_{i1} + J_{i2}) = \rho_1 - \rho_2 + \rho_2 - \rho_3 = \rho_1 - \rho_3$) and the fact that $\|\|$ itself satisfies the triangle inequality. The same considerations now guarantee that the discrete approximations of (2) are also distance functions.

Alternative form for biharmonic distance  Take $G(\cdot, \cdot)$ to be the Green’s function of the Laplace-Beltrami operator $\Delta$, with eigenfunctions $\phi_i$ and eigenvalues $\lambda_i$ for $i = 1, 2, \ldots$. As it is defined in [Lipman et al. 2010], the biharmonic distance between $p, q \in M$ is given by
d

$$d_b(p, q) = \sum_i \left( \frac{\phi_i(p) - \phi_i(q)}{\lambda_i} \right)^2 \int M \phi_i(x) \phi_i(x) dx$$

Distributing the difference in the integrand and using the relationship $G(x, p) = \sum_i \frac{\phi_i(x) \phi_i(p)}{\lambda_i}$ shows $d_b(p, q) = \int M (G(x, p) - G(x, q))^2 dx$, as desired.

Proof that volumetric distance is bounded below by Euclidean distance  The barycentric coordinates of a point $p \in \mathbb{R}^3$ is a probability distribution that we will denote in this proof by $w_p \in \text{Prob}(M)$. It satisfies $\int_M w_p(x) dx = 1$ (partition of unity property) and $\int_M x w_p(x) dx = p$ (linear precision property). Suppose now that $p, q \in \mathbb{R}^3$ have barycentric coordinates $w_p, w_q$. Let $J$ be the optimal vector field for the optimization problem (2) and let $r(x) \equiv \|x\| \cdot \text{sgn}(w_p(x) - w_q(x))$ be a “signed” distance to the origin in $\mathbb{R}^3$, satisfying $\|\nabla r(x)\| = 1$ almost everywhere. Then

$$d_V(p, q) = \int_M \|J\| dx \geq \int M \|\nabla r\| dx$$

by the triangle inequality. We now apply Stokes’ Theorem on $M$ and invoke the divergence constraint satisfied by $J$. This yields

$$d_V(p, q) \geq \int M r \cdot J dx = \int M \|\nabla r\| w_p(x) - w_q(x) dx \geq \|w_p(x) - w_q(x) dx\| = ||p - q||,$$

by the triangle inequality and the use of barycentric coordinates.